

UK Patent Application

(19) GB (11) 2 212 511 (13) A

(43) Date of A publication 26.07.1988

(21) Application No 8827024.4

(22) Date of filing 18.11.1988

(30) Priority data

(31) 122714 (32) 18.11.1987 (33) US
139886 30.12.1987
161072 26.02.1988
191263 06.05.1988
263594 26.10.1988

(51) INT CL*

C12N 7/00, A61K 39/29 39/42, C12N 1/00 15/00, C12P
G01N 33/53 //
(C12N 15/00 C12P 21/00 C12R 1:91)

(52) UK CL (Edition J)

C6F FJ F10X
C3H HB4 HB7P H650 H654 H655 H674 H690
C6Y Y125 Y330 Y344 Y403 Y501 Y503
U1S S1332 S1337 S2419

(71) Applicant

Chiron Corporation

(Incorporated in the USA - Delaware)

4560 Horton Street, Emeryville, California 94608,
United States of America

(72) Inventors

Michael Houghton
Qui-Lim Choo
George Kuo

(74) Agent and/or Address for Service

J A Kemp and Co
14 South Square, Gray's Inn, London, WC1R 5EU,
United Kingdom

(56) Documents cited

EP 0277437 A2 WO 88/03410 A1 WO 82/02774 A1
WO 82/00205 A1 US 4895976 A
Naturwissenschaften 1986, 73(1)45-46
Gastroenterology 1979, 77(5), A33 (30th Annual
Meeting of the American Association for the study
of liver disease, Chicago, Ill., USA, Nov. 7, 1979)
Postgrad. Med. USA 1980, 68(5), 66-86
Immun. Infect. 1988, 16(3), 97-99
Immun. Infect. 1988, 16(3), 91-96
Chiron News Release 1988, May 10, 21
Lancet 1978, 21 October, 8095

(58) Field of search

UK CL (Edition J) C3H HB7P, C6F FJ
INT CL* C12N 7/00 15/00
ON-LINE 'Dialog' (biotech, Derwent WPI)

(54) Hepatitis C virus

(57) A family of cDNA sequences derived from hepatitis C virus (HCV) are provided. These sequences encode antigens which react immunologically with antibodies present in individuals with non-A non-B hepatitis (NANBH), but which generally are absent from individuals infected with hepatitis A virus (HAV) or hepatitis B virus (HBV), and also are absent from control individuals. A comparison of these cDNA sequences with the sequences in Genebank, and with the sequences of hepatitis delta virus (HDV) and HBV shows a lack of substantial homology. A comparison of the sequences of amino acids encoded in the cDNA with the sequences of Flaviviruses indicates that HCV is a Flavivirus or Flavi-like virus.

The HCV cDNA sequences are useful for the design of polynucleotide probes, and for the synthesis of polypeptides which may be used in immunoassays. Both the polynucleotide probes and the polypeptides may be useful for the diagnosis of HCV-induced NANBH, and for screening blood bank specimens and donors for HCV infections. In addition, these cDNA sequences may be useful for the synthesis of immunogenic polypeptides which may be used in vaccines for the treatment, prophylactic and/or therapeutic, of HCV infection. Polypeptides encoded within the cDNA sequences may also be used to raise antibodies against HCV antigens, and for the purification of antibodies directed against HCV antigens. These antibodies may be useful in immunoassays for detecting HCV antigens associated with NANBH in individuals, and in blood bank donations. Moreover, these antibodies may be used for treatment of NANBH in individuals.

The reagents provided in the invention also enable the isolation of NANBH agent(s), and the propagation of these agent(s) in tissue culture systems. Moreover, they provide reagents which are useful for screening for antiviral agents for HCV, particularly in tissue culture or animal model systems.

GB 2 212 511 A

2212511

FIG. 1 Translation of DNA 5-1-1

1 AlaSerCysLeuAsnCysSerAlaSerIleIleProAspArgGluValLeuTyrArgGlu
GGCCCTCCTGCTTGAACTGCTCGCGAGCATACCTGACAGGGAAAGTCCTTACCGAGA
CCGGAGGACGAACTTGACGAGCCGCTCGTAGTATGGACTGTCCCTCAGGAGATGGCTCT
PheAspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeu
61 GTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCT
CAAGCTACTCTACCTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTCCCTACTACGA
AlaGluGlnPheLysGlnLysAlaLeuGlyLeu
121 CGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCCTCC
GCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGG

FIG. 3 Translation of DNA 5-1-1, 81, 91&1-2

1 GlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIleIleProAsp
CTGGCTCGTGTCATAGTGGGCAGGGTCGTCTGTCCGGGAAGCCGGCAATCATACCTG
GACCGACGCACCAGTATCACCGTCCAGCAGAACAGGCCCTCGGCCGTAGTATGGAC
T
61 ArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyr
ACAGGGAAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGT
TGTCCCTTCAGGAGATGGCTCTCAAGCTACTCTACGAGAGTCGTGAATGGCA
A
121 IleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeuLeuGln
ACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAACAGGCCCTCGGCCCTGC
TGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTCCGGGAGCCGGAGGACG
ThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGlnLysLeu
181 AGACCGCGTCCCGTCAGGCAGAGGTATCGCCCTGCTGTCCAGACCAACTGGCAAAAC
TCTGGCGCAGGGCAGTCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTTTIG
GluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeuAlaGly
241 TCGAGACCTCTGGGCAGACATATGGAACCTTCATCAGTGGATAACAATACTGGCGG
AGCTCTGGAAGACCGCTCGTATACACCTTGAAAGTAGTCACCCTATGTTATGAACCGCC
LeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAlaAlaVal
301 GCTTGTCAACGCTGCCTGTTAACCCGCCATTGCTTCATTGATGGCTTTACAGCTGCTG
CGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACCTACCGAAAATGTCGACGAC
ThrSerProLeuThrThrSerGln
361 TCACCAGCCCCACTAACCACTAGCCAAA
AGTGGTCGGGTATTGGTATCGGTT

2212511

5-1-1	1		GGCTCTGTTGAACTGCTGGAGAGCATTGAGATCGAAGAGTGTCTCAGGACTTACCGTACATCGAGCAAGGGATGATGC
81	1		GTCGGGAAGCGGCCAATCATACCTGACAGGGAG
91	1	ctggctgttgttGTCATACTGGGGCAAGGGCGCTTGTCCGGAAAGCGGCCAATCATACCTGACAGGGAG	
1-2	1		GGTCATACTGGCCAAGGGCGCTTGTCCGGAAAGCGGCCAATCATACCTGACAGGGAG
5-1-1	48	TCCCTCTACCGAGAGTTCGATGAGATCGAAGAGTGTCTCAGGACTTACCGTACATCGAGCAAGGGATGATGC	
81	36	TCCCTCTACCGAGAGTTCGATGAGATCGAAGAGTGTCTCAGGACTTACCGTACATCGAGCAAGGGATGATGC	
91	70	TCCCTCTACCGAGAGTTCGATGAGATCGAAGAGTGTCTCAGGACTTACCGTACATCGAGCAAGGGATGATGC	
1-2	60	TCCCTCTACCGAGAGTTCGATGAGATCGAAGAGTGTCTCAGGACTTACCGTACATCGAGCAAGGGATGATGC	
5-1-1	120	TCCCCGAGGCAAGTCAAGCAGAACGCCCTCGGCCCTCC	
81	108	TCGGCGAGCAAGTCAAGCAGAACGCCCTCGGCCCTCC	
91	142	TCGGCGAGCAAGTCAAGCAGAACGCCCTCGGCCCTCC	
1-2	132	TCGGCGAGCAAGTCAAGCAGAACGCCCTCGGCC	
81	180	CTGGCTTGCCAGACCAACTTGGCAAAAAGTGGACCTTCTGGCGAAGCCATTGTTGAACTTCATCTGGAACTTCATCTGGCA	
91	214	CTGGCTTGCCAGACCAACTTGGCAAAAAGTGGACCTTCTGGCGAAGCCATTGTTGAACTTCATCTGGAACTTCATCTGGCA	
81	252	TACAACTACTTGGGGCTTGTCAACGGTGCCTCTGAACTGGCTTGTCAACGGTGCCTGG	
91	286	TACAACTACTTGGGGCTTGTCAACGGTGCCTGG	

2
EIG

221251

3

FIG. 4 Translation of DNA 81

SerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMet
1 GTCCGGGAAGCCGGCAATCATACCTGACAGGGAACTCCTACCGAGAGTCGATGAGAT
CAGGCCCTCGGCCGTAGTATGGACTGTCCTCAGGAGATGGCTCTCAAGCTACTCTA

GluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPhe
61 GGAAGAGTGCTCTCACGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTT
CCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAA

LysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaPro
121 CAAGCAGAAGGCCCTCGGCCTCCTGCAGACCAGCAGTCAGGAGGTTATGCC
GTTCGTCTCCGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCAAATAGCGGGG

AlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPhe
181 TGCTGTCCAGACCAACTGGCAAAACTCGAGACCTCTGGCGAACGATATGTGGAAC
ACGACAGGTCTGGTTGACCGTTTGAGCTCTGGAAAGACCCGCTCGTATACACCTTGAA

IleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAla
241 CATCAGTGGGATAACAATACTTGGCGGGCTTGTCAACGCTGCCGTGTAACCCGCCATTGC
GTAGTCACCCATATGTTATGAACGCCAACAGTGCACGGACCATTGGGCGGTAAACG

SerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
301 TTCATTGATGGCTTTACAGCTGCTGCAACCAGCCACTAACCACTAGCCAAA
AAGTAACCTACCGAAAATGTCGACGACAGTGGTCGGGTATTGGTATCGGTTT

FIG. 5 Translation of DNA 36

AspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAla
1 GATGCCCACTTCTATCCCAGACAAAGCAGAGTGGGAGAACCTCCTACCTGGTAGCG
CTACGGGTGAAAGATAGGGCTGTTCTCACCCCTCTGGAAAGGAATGGACCATCGC

TyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrp
61 TACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCATCGTGGGACCAGATGTGG
ATGGTTGGTGGCACACGCGATCCCGAGTTGGGGAGGGTAGCACCCTGGTCTACACC

LysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu
121 AAGTGGTTGATTGCCCTCAAGCCCACCCCTCCATGGCCAACACCCCTGCTATACAGACTG
TTCACAAACTAACGGAGTTGGGTGGGAGGTACCCGGTTGTGGGGACGATATGTCGAC

GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCys
181 GGCCTGTTAGAAATCACCTGACGCACCCAGTCACCAAATACATCATGACATGC
CCGCGACAAGTCTTACTTTAGTGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACG

MetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAla
241 ATGTCGGCCGACCTGGAGGTGTCACGAGCACCTGGGTGCTCGTGGCGCGTCTGGCT
TACAGCCGGCTGGACCTCCAGCAGTGCCTGGACCCACGAGCAACCGCCGAGGACCGA

AlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeu
301 GCTTTGGCCCGTATTGCTGTCAACAGGCTGCGTGGTCAGTGGCAGGGTCGTCTTG
CGAAACCGGGCGATAACGGACAGTTGTCGACGCACCGATATCACCCGTCCCAGCAGAAC

-----Overlap with 81-----
SerGlyLysProAlaIleIleProAspArgGluValLeuTyrArg
361 TCCGGGAAGCCGGCAATCATACCTGACAGGGAACTCCTACCGAG
AGGCCCTCGGCCGTAGTATGGACTGTCCTCAGGAGATGGCTC

221251

4

FIG. 6 Combined ORF of DNAs 36 & 81

AspAlaHisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAla
1 GATGCCCACTTCTATCCAGACAAAGCAGTGAGGGAGAACCTTACCTGGTAGCG
CTACGGGTGAAAGATAGGGTCTGTTCTCACCCCTCTTGGAAAGGAATGGACCATCGC
TyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrp
61 TACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCATCGTGGGACCAGATGTGG
ATGGTTGGTGGCACACGCGATCCCGAGTTCTGGGGAGGGGTAGCACCCCTGGTCTACACC
LysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu
121 AAGTGTGATTGCCTCAAGCCCACCCCATGGCCAACACCCCTGCTATACAGACTG
TTCACAAACTAAGCGGAGTTGGGTGGGAGGTACCCGGTTGTGGGACGATATGCTGAC
GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCys
181 GGCCTGTTAGAAATCACCTGACGCACCCAGTCACCAAATACATCATGACATGC
CCGCGACAAGTCTTACTTTAGGGACTGCGTGGTCAGTGGTTATGTAGTACTGTACG
MetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAla
241 ATGTCGGCCGACCTGGAGGTGTCAGCAGCACCTGGGTGCTCGTGGCGGCGTCTGGCT
TACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACGCCAGGACCGA
AlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeu
301 GCTTGGCCGCGTATTGCCTGTCACAGGCTGCGTGGTCATAAGGGCAGGGTCGTCTTG
CGAAACCGGCGATAACGGACAGTTGTCCGACGCACAGTATCACCGTCCCAGCAGAAC
SerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMet
361 TCCGGGAAGCCGGCAATCATACCTGACAGGGAAGTCCTCTACCGAGAGTCGATGAGATG
AGGCCCTCGGCCGTTAGTATGGACTGTCCTCAGGAGATGGCTCTCAAGCTACTCTAC
GluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPhe
421 GAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTGCCGAGCAGTTC
CTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTCCCTACTACGAGCCGCTCGTCAAG
LysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaPro
481 AAGCAGAAGGCCCTCGGCCCTCGAGACCGCGTCCCGTCAGGCAGAGGTTATCGCCCT
TTCGTCTCCGGGAGCCGGAGGACGTCGGCGCAGGGCAGTCGTCTCCAATAGCGGGGA
AlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPhe
541 GCTGTCCAGACCAACTGGAAAAACTCGAGACCTTCTGGGCGAACGCATATGTGAACTTC
CGACAGGTCTGGTTGACCGTTTGAGCTCTGGAAAGACCCGCTCGTATACACCTGAAG
IleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAla
601 ATCAGTGGGATAACAATCTGGCGGGCTTGTCAACGCTGCCTGGTAACCCGCCATTGCT
TAGTCACCCATGTTATGAACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTACCGA
SerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
661 TCATTGATGGCTTTACAGCTGCTGTCACCAGCCACTAACCACTAGCCAAA
AGTAACATCCGAAAATGTCGACGACAGTGGTCGGGTATTGGTGATCGGTTT

2212511

FIG. 7 Translation of DNA 32

-----Overlap with 81-----
PheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeu
1 CTTTTACAGCTGCTGTACCAGCCCCACTAACCACTAGCCAAACCCCTCTTCACATAT
GAAAATGTCGACGACAGTGGTCGGGTGATTGGTATCGGTTGGGAGGAGAAGTTGTATA

GlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAla
61 TGGGGGGGTGGGTGGCTGCCAGCTCGCCGCCGGTGCCGCTACTGCCTTGCTGGCG
ACCCCCCACCACCGACGGGTCGAGCGGCCACGGCGATGACGAAACACCCGC.

GlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeu
121 CTGGCTTAGCTGGCGCCCATCGGCAGTGTGGACTGGGAAGGTCTCATAGACATCC
GACCGAACATCGACCGCGCGGTAGCCGTACAACCTGACCCCTCCAGGAGTATCTGTAGG

AlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGlu
181 TTGCAGGGTATGGCGCGGGCGTGGCGGGAGCTTGTGGCATCAAGATCATGAGCGGTG
AACGTCCCATAACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCAC

ValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeu
241 AGGTCCCCCTCCACGGAGGACCTGGTCAATCTACTGCCGCCATCCTCTGCCCGGAGCCC
TCCAGGGGAGGTGCCCTCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGGCCTCGGG

ValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAla
301 TCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCCGCCACGTTGGCCCGGGCGAGGGGG
AGCATCAGCCGCACCAAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCGCTCCCC

ValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
361 CAGTGCAGTGGATGAACCGGGCTGATAGCCTCGCCCTCCGGGGAACCATGTTCCCC
GTCACGTCACCTACTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGG

221251

FIG. 8 Translation of DNA 35

6

SerIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
1 TCCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCGCACTCACGTCGGGGCAGG
AGGTAACTCTGTTAGTGCAGGGGGCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCC

ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
61 ACTGGCAGGGGGAAAGCCAGGCATCTACAGATTGTGGCACCGGGGGAGCGCCCTCCGGC
TGACCGTCCCCCTCGGTCCGTAGATGTCTAACACCGTGGCCCCCTCGCGGGGAGGCCG

MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
121 ATGTTGACTCGTCCGTCTGTGAGTGTCTATGACGCAGGCTGTGCTTGTATGAGCTC
TACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAG

ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
181 ACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCGGGGCTTCCCCTG
TGCAGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCAC

CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
241 TGCCAGGACCATCTGAATTTGGGAGGGCGTCTTACAGGCCTACTCATATAGATGCC
ACGGTCCTGGTAGAACTTAAACCCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGG

HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
301 CACTTTCTATCCAGACAAAGCAGAGTGGGAGAACCTTCCTTACCTGGTAGCGTACCAA
GTGAAAGATAGGGCTGTTCGTCTCACCCCTTTGGAAAGGAATGGACCATCGCATGGTT

-----Overlap with 36-----

AlaThrValCysAlaArgAlaGinAlaProProProSerTrpAspGlnMetTrpLysCys
361 GCCACCGTGCGCTAGGGCTCAAGCCCCCTCCCCCATCGTGGGACCAGATGTGGAAGTGT
CGGTGGCACACGCGATCCCGAGTTCGGGGAGGGGGTAGCACCCTGGTCTACACCTTCACA

LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
421 TTGATT CGCCTCAAGCCCACCCCTCCATGGGCCAACACCCCTGCTATACAGACTGGCGCT
AACTAAGCGGAGTTCGGGTGGGAGGTACCCGGTTGGGGACGATATGTCTGACCCCGCA

FIG. 9-1 Combined ORF of DNAs 35, 36, 81 & 32

1 SerIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
 TCCATTGAGACAATCACGCTCCCCCAGGATGCTGTCTCCGCACCAACGTCGGGCAGG
 AGGTAACCTGTAGTGCAGGGGGCTTACGACAGAGGGCGTAGTTGCAGCCCCGTCC

 61 ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
 ACTGGCAGGGGAAGCCAGGCATCTACAGATTGTGGCACCGGGGAGCGCCCTCCGGC
 TGACCGTCCCCCTCGTCCGTAGATGTCTAACACCGTGGCCCCTCGCGGGGAGGCCG

 121 MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
 ATGTTCGACTCGTCCGTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTC
 TACAAGCTGAGCAGGCAGGAGACACTCACGATACTCGTCCGACACGAACCATACTCGAG

 181 ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
 ACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGCTTCCGTG
 TGGGGCGGCTCTGATGTCAATCGATGCTCGATGTACTTGTGGGCCCGAAGGGCAC

 241 CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
 TGCCAGGACCATCTGAATTGGGAGGGCGTCTTACAGGCTCACTCATATAGATGCC
 ACGGTCTGGTAGAACTTAAACCTCCCGAGAAATGTCCGGAGTGAGTATATCTACGG

 301 HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
 CACTTCTATCCAGACAAAGCAGAGTGGGAGAACCTCCTACCTGGTAGCGTACCAA
 GTGAAAGATAGGGTCTGTTCGTCTCACCCCTCTTGAAGGAATGGACCATCGCATGGTT

 361 AlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCys
 GCCACCGTGTGCGCTAGGGCTCAAGCCCCCTCCCCATCGTGGGACCAGATGTGGAAGTGT
 CGTGGCACACCGCATTGGAGTTCGGGAGGGTAGCACCTGGTCTACACCTTCACA

 421 LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
 TTGATTGCGCTCAAGCCCACCCCTCCATGGGCAACACCCCTGCTATACAGACTGGCGCT
 AACTAAGCGGAGTTCGGTGGGAGGTACCCGGTTGTGGGACGATATGTCTGACCCGCGA

 481 ValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSer
 GTTCAGAATGAAATCACCTGACGCACCCAGTCACCAAATACATCATGACATGCATGTCG
 CAAGCTTACTTTAGTGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACAGC

 541 AlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeu
 GCCGACCTGGAGGTGTCACGAGCACCTGGGTGCTCGTTGGCGCGTCTGGCTGCTTG
 CGGCTGGACCTCCAGCAGTGCCTGGACCCACGAGCAACGCCCGCAGGACCGACGAAAC

 601 AlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeuSerGly
 GCCGCGTATTGCGCTGTCAACAGGCTGCGTGGTCATAGTGGCAGGGTCGTCTGTCGGG
 CGCGCATAACGGACAGTTGTCAGCACCAGTACCCGTCAGCAGAACAGGCCCC

 661 LysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGlu
 AAGCCGGCAATCATACCTGACAGGGAAAGTCCTCTACCGAGAGTTCGATGAGATGGAAGAG
 TTCGGCCGTTAGTGGACTGTCCTTCAGGAGATGGCTCTCAAGCTACTCACCTTC

 721 CysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGln
 TGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAG
 ACGAGAGTCGTGAATGGCATGTAGCTCGTCCCTACTACGAGCGGCTCGTCAAGTCGTC

 781 LysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaVal
 AAGGCCCTCGGCCTCCTGCAGACCGCGTCCCCGTCAAGCAGAGGTTATGCCCTGCTGTC
 TTCCGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGGACGACAG

2212511

4

GlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSer
841 CAGACCAACTGGCAAAACTCGAGACCTTCTGGGCGAAGCATAATGTGGAACCTCATCAGT
GTCTGGTTACCGTTTGAGCTCTGAAAGACCCGCTCGTATACACCTTGAAGTAGTCA

GlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeu
901 GGGATACAAATACTTGGCGGGCTTGTCAACGCTGCCCTGGTAACCCGCCATTGCTTCATTG
CCCTATGTATGAACCGCCCCAACAGTGGCACGGACCATTGGGCGGTAACGAAGTAAC

MetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsn
961 ATGGCTTTACAGCTGCTGTCAACCAGCCCCTAAACCAACTAGCCAACCCCTCCTTCAC
TACCGAAAATGTCGACGACAGTGGTCGGGTATTGGTATCGGTTGGGAGGAGAACATTG

IleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheVal
1021 ATATTGGGGGGGTGGGTGGCTGCCAGCTGCCGCCGGTGCCGCTACTGCCTTG
TATAACCCCCCCCACCCACCGACGGGTCGAGCGGGCGGGGCCACGGCGATGACGGAAACAC

GlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAsp
1081 GGCCTGGCTTAGCTGGCGCCCATCGCAGTGGACTGGGAAGGTCTCATAGAC
CCGCGACCGAACATCGACCGCGGGTAGCCGTACAACCTGACCCCTCCAGGAGTATCTG

IleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSer
1141 ATCCTTGAGGGTATGGCGCGGGCTGGCGGGAGCTCTTGCGATTCAAGATCATGAGC
TAGGAACGTCCCATAACCGCGCCCGCACGCCCTCGAGAACACCGTAAGTTCTAGTACTCG

GlyGluValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGly
1201 GGTGAGGTCCCCCTCCACGGAGGACCTGGTCAATCTACTGCCGCCATCCTCTGCCCGGA
CCACTCCAGGGAGGTGCCTCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGGCC

AlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGlu
1261 GCCCTCGTAGTCGGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAG
CGGGAGCATCAGCCGACCGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCGCTC

GlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
1321 GGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTTCGCCTCCGGGGAAACCATGTTCCCC
CCCCGTACGTCACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGG

FIG. 9-2

FIG. 10

Translation of DNA 37b

Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp
 1 CTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCCGTGGCCTACTACCGCGGTCTTGAC
 GAGCGGCAGTTCGACCAGCGTAACCGTAGTTACGGCACCGGATGATGGCGCCAGAACGT

Val Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
 61 GTGTCCGTCATCCCGACCAGCGCGATGTTGTCGTCGTGGCAACCGATGCCCTCATGACC
 CACAGGCAGTAGGGCTGGTCGCCGCTAACACAGCAGCACCGTGGCTACGGGAGTACTGG

Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Tyr Asn Thr Cys Val Thr Gln Thr Val
 121 GGCTATAACCGGCGACTTCGACTCGGTAGACTACAATACTGTTGTCACCCAGACAGTC
 CCGATATGGCCGCTGAAGCTGAGCCACTATCTGATGTTATGCACACAGTGGGTCTGTCAG

-----Overlap with-----
 181 Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp Alanyl
 GATTTCAGCCTTGACCCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGAAGCTGTC
 CTAAAGTCGGAACCTGGGATGGAAGTGGTAACCTGTTAGTGCAGGGGGTCTACGACAG

clone 35-----
 Ser Arg Thr Gln Arg Arg Gly Arg Thr
 241 TCCCGCACTCAACGTCGGGGCAGGACTG
 AGGGCGTGAGTTGCAGCCCCGTCCTGAC

FIG. 11

Translation of DNA 33b

-----Overlap with 32-----
 1 Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val
 GATGAACCGGCTGATAGCCTTCGCCTCCGGGGAACCATGTTCCCCCACGCACACTACGT
 CTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGGGTGCCTGATGCA

61 Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln
 GCCGGAGAGCGATGCGAGCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCA
 CGGCCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTCGGAGTGACATTGGGT

121 Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser
 GCTCCTGAGGCAGTCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTC
 CGAGGACTCGCTGACGTGGTCACCTATTGAGCCTCACATGGTGAGGTACGAGGCCAG

181 Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu
 CTGGCTAAGGGACATCTGGACTGGATATGCGAGGTGTTGAGCGACTTTAACGACCTGGCT
 GACCGATTCCCTGTAGACCCCTGACCTATACTGCTCCACAACGCTGAAATTCTGGACCGA

241 Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr
 AAAAGCTAACGCTCATGCCACAGCTGCCCTGGGATCCCTTTGTGTCCTGCCAGCGCGGGTA
 TTTTCGATTGAGTACGGTGTGACGGACCCCTAGGGGAAACACAGGACGGTACGCGCCCCAT

301 Lys Gly Val Trp Arg Val
 TAAGGGGGTCTGGCGAGTG
 ATTCCCCCAGACCGCTCAC

FIG. 12 Translation of DNA 40b

AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
 1 GGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCGGGTGAGAACAAAT
 CGAATGTACAGGTTCCGAGTAGCCTAGCTAGGATTGATGCTCTGGCCCCACTCTTGTAA

 ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
 61 TACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCCTGCCGACGGCGGGTG
 ATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCCTCAAGGAACGGCTGCCGCCAC

 SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
 121 CTCGGGGGGCGCTTATGACATAATAATTGTGACGAGTGCACCCACGGATGCCACATC
 GAGCCCCCGCGAATACTGTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGAG

 IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
 181 CATCTTGGGCATCGGCACTGTCCTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTGT
 GTAGAACCCGTAGCCGTGACAGGAACCTGGTCTCTGACGCCCGCTTGACCAACA

 LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
 241 GCTCGCCACCAGCCACCCCTCCGGGCTCCGTCACTGTGCCCATCCAAACATCGAGGAGGT
 CGAGCGGTGGCGGTGGGGAGGCCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCCTCA

 AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
 301 TGCTCTGTCCACCACCGGAGAGATCCCTTTACGGCAAGGCTATCCCCCTCGAAGTAAT
 ACGAGACAGGTGGTGGCCTCTAGGGAAAAATGCCCTCGATAGGGGAGCTTCATTA

 LysGlyGlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAla
 361 CAAGGGGGGGAGACATCTCATCTCTGTATTCAAAGAAGTGCACGAACTCGCCGC
 GTTCCCCCCCCTCTGTAGAGTAGAACAGTAAGTTCTTCACGCTGCTIGAGCGGGCG

 -----Overlap with 37b-----
 LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
 421 AAAGCTGGTCGCATTGGGCATCAATGCCGTGGCTACTACCGCGGTCTGACGTGTCCGT
 TTTCGACCAAGCGTAACCGTAGTTACGGCACCGGATGATGGCGCCAGAACTGCACAGGCA

 IleProThr
 481 CATCCGACCAAG
 GTAGGGCTGGTC

2212511

11

FIG. 13 Translation of DNA 25c

CysSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCys
1 ACTGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGT
TGACGTCGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGCACCTATTCGAGCCTCA

ThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeu
61 GTACCACTCCATGCTCCGGTTCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGT
CATGGTGAGGTACGAGGCCAAGGACCGATTCCCTTAGACCCGTACCTATACGCTCCACA

-----Overlap with 33b-----
SerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGlyIleProPhe
121 TGAGCGACTTTAACGACCTGGCTAAAAGCTAACGCTCATGCCACAGCTGCCTGGATCCCCT
ACTCGCTGAAATTCTGGACCGATTTCGATTGAGTACGGTGTGACGGACCCCTAGGGGA

ValSerCysGlnArgGlyTyrLysGlyValTrpArgGlyAspGlyIleMetHisThrArg
181 TTGTGTCTGCCAGCGCGGGTATAAGGGGTCTGGCGAGGGGACGGCATCATGCACACTC
AACACAGGACGGTCGCCCCATATTCCCCCAGACCGCTCCCTGCCGTAGTACGTGTGAG

CysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArgIleValGly
241 GCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAACGGGACGAIGAGGATCGTCG
CGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTTGCCCTGCTACTCCTAGCAGC

ProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGly
301 GTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTCCCCATTAATGCCCTACACCACGG
CAGGATCCTGGACGTCCCTGTACACCTCACCTGGAAAGGGGTATTACGGATGTGGTGC

ProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGlu
361 GCCCCCTGTACCCCCCTTCCTGCCCGAACATCACGTTGCGCTATGGAGGGTGTGCAAG
CGGGGACATGGGGGAAGGACGCCGCTTGATGTGCAAGCGCGAIACCTCCCACAGACGTC

GluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAsp
421 AGGAATATGTGGAGATAAGGCAGGTGGGGACTTCACTACGTGACGGGTATGACTACTG
TCCTTATACACCTCTATICCCTCACCCCTGAAGGTATGCAACTGCCATACTGATGAC

AsnLeuLysCysProCysGlnValProSerProGluPhePheThrGlu
481 ACAATCTCAAATGCCGTGCCAGGTCCCATGCCCGAATTTCACAGAAAT
TGTTAGAGTTACGGGCACGGTCCAGGGTAGCGGGCTTAAAAGTGTCTTA

2212511

FIG. 14-1 Combined ORF of DNAs 40b/37b/35/36/81/32/33b/25c

1 AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
 1 TGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCGGGTGAGAACAAAT
 ACGAATGTACAGGTCCGAGTAGCCTAGTAGGATTGTAGTCCTGGCCCCACTCTGTAA
 61 ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
 61 TACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCCTGCCGACGGCGGGTG
 ATGGTGACCGTCGGGTAGTCATGAGGTGGATGCCGTTAAGGAACGGCTGCCGCCAC
 121 SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
 121 CTCGGGGGCGCTTATGACATAATAATTGTGACGAGTGCACACTCCACGGATGCCACATC
 GAGCCCCCGCGAATACTGTATTATAACACTGCTCACGGTGAGGTGCCTACGGTGTAG
 181 IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
 181 CATCTGGGCATCGGCACTGTCCTTGACCAAGCAGAGACTGCCGGGCGAGACTGGTTGT
 GTAGAACCCGTAGCCGTGACAGGAACCTGGTCGTCAGCAGCCCCGCTCTGACCAACA
 241 LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
 241 GCTCGCCACCGCCACCCCTCCGGCTCGTCACTGTGCCCATCCAACATCGAGGAGGT
 CGAGCGGTGGCGGTGGGGAGGCCCAGGCAGTGACACGGGTAGGGTTGTAGCTCCTCCA
 301 AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
 301 TGCTCTGCAACCAACCGGAGAGATCCCTTTACGGCAAGGCTATCCCCCTCGAAGTAAT
 ACGAGACAGGTGGTGGCCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTICATTIA
 361 LysGlyGlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAla
 361 CAAGGGGGGAGACATCTCATCTCTGTCATCAAAGAAGAAGTGCACGAACTCGCCGC
 GTTCCCCCTCTGAGACTAGAAGACAGTAAGTTCTTACGCTGCTTGACGGCG
 421 LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
 421 AAAGCTGGTCGCATTGGGCATCAATGCCGTGGCTACTACCGCGTCTGACGTGCGT
 TTCGACCAAGCGTAACCCGTAGTTACGGCACCGGATGATGGGCCAGAACCTGCACAGGC
 481 IleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGlyTyrThr
 481 CATCCCGACCAGCGCGATGTTGTCGTGTCGGCAACCGATGCCCTCATGACCGGTATAC
 GTAGGGCTGGTCGCCGCTACACAGCAGCACCGTGGCTACGGGAGTACTGGCGATATG
 541 GlyAspPheAspSerValIleAspTyrAsnThrCysValThrGlnThrValAspPheSer
 541 CGGCGACTTCGACTCGGTGATAGACTACAATACGTTGTCACCCAGACAGTCGATTTCAG
 GCCGCTGAAGCTGAGCCACTATCTGATGTTATGCACACAGTGGGTCTGTCAGCTAAAGTC
 601 LeuAspProThrPheThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThr
 601 CCTTGACCCCTACCTCACCATGAGACAATCACGCTCCCCCAGGATGCTGTCCTCCGCAC
 GGAACTGGGATGGAAGTGGTAACCTGTTAGTGCAGGGGGTCTACGACAGAGGGCGTG
 661 GlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGly
 661 TCAACGTCGGGGCAGGACTGGCAGGGGAAGCCAGGCATCTACAGATTGTGGCACCGGG
 AGTTGCAGCCCCGTCCTGACCGTCCCCCTCGGTCCGTAGATGTCTAAACACCGTGGCCC
 721 GluArgProSerGlyMetPheAspSerSerValIleCysGluCysTyrAspAlaGlyCys
 721 GGAGCGCCCTCCGGCATGTTGACTCGTCCCTGTGAGTGCTATGACGGCAGGCTG
 CCTCGGGGGAGGCCGTACAAGCTGAGCAGGCAGGAGACACTCACGATACTGCGTCCGAC
 781 AlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThr
 781 TGCTTGGTATGAGCTACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAAACAC
 ACGAACCATACTCGAGTGCAGGGCGGCTCTGATGTCATGCTCGCATGTACTTGTG
 841 ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeu
 841 CCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAATTGGGAGGGCGTCTTACAGGCCT
 GGGCCCCGAAGGGCACACGGCCTGGTAGAACTTAAACCCCTCCCGCAGAAATGTCCGGA
 901 ThrHisIleAspAlaHisPheLeuSerGlnSerGlyGluAsnLeuProTyr
 901 CACTCATATAGATGCCACTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTCCTTA
 GTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTCGTCTACCCCTCTTGAAGGAAT

961 LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAsp
 CCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCTCCCCATCGTGGGA
 GGACCATCGCATGGTTCGGTGGCACACCGATCCCAGTTGGGGAGGGTAGCACCCT

1021 GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu
 CCAGATGTGGAAGTGTGATTGCCCTCAAGCCCACCCCTCCATGGGCCAACACCCCTGCT
 GGTCTACACCTCACAAACTAACCGGAGTTGGGTGGGAGGTACCCGGTTGTGGGGACGA

1081 TyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle
 ATACAGACTGGGCCTGTTCAAGAATCACCCCTGACGCACCCAGTCACCAAATACAT
 TATGTCGACCCCGCACAAGTCTTACTTTAGTGGGACTGCGTGGTCAGTGGTTATGTA

1141 MetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly
 CATGACATGCATGTCGGCCGACCTGGAGGTGCGTACCGAGCACCTGGGTGCTCGTTGGCGG
 GTACTGTACGTACAGCCGGCTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGC

1201 ValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArg
 CGTCCTGGCTGCTTGGCCCGCTATTGCCGTCAACAGGCTGCGTGGTCATAGTGGGCAG
 GCAGGACCGACGAAACCGGCCATAACGGACAGTTGTCGACGCACCAAGTATCACCCGTC

1261 ValValLeuSerGlyLysProAlaIleIleProAspArgGluValLeuTyrArgGluPhe
 GGTCGTCTGTCCGGGAAGCCGGCAATCATACCTGACAGGGAAAGTCCTCTACCGAGAGTT
 CCAGCAGAACAGGCCCTCGGCCGTTAGTATGGACTGTCCTTCAGGAGATGGCTCTCAA

1321 AspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAla
 CGATGAGATGGAAGAGTGCTCTCAGCACTACCGTACATCGAGCAAGGGATGATGCTCGC
 GCTACTCTACCTCTCACGAGAGTCGTGAATGGCATGTCAGTCCTACTACGAGCG

1381 GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluVal
 CGAGCAGTTCAAGCAGAAGGCCCTCGGCCCTGCAGACCGCGTCCCGTCAGGCAGAGGT
 GCTCGTCAAGTTCGTCTCCGGGAGCCGGAGCTGTCGGCGCAGGGCAGTCCTCA

1441 IleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMet
 TATCGCCCTGCTGTCCAGACCAACTGGCAAAAACTCGAGACCTCTGGGCGAAGCATAT
 ATAGCGGGGACGACAGGTCTGGTTGACCGTTTGAGCTCTGGAAAGACCCGTTCGTATA

1501 TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro
 GTGGAACTTCATCAGTGGATACAATACTGGCGGGCTGTCAACGCTGCCCTGGTAACCC
 CACCTTGAAAGTAGTCACCCATTGTTATGAACCGCCCAACAGTTGCGACGGACCATGGG

1561 AlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
 CGCCATTGCTTCATTGATGGCTTTACAGCTGCTGTCAACCAGCCCCTAACCAACTAGCCA
 CGGGTAACGAAGTAACCTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGGT

1621 ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAla
 AACCCCTCCTCTTCAACATATTGGGGGGTGGCTGCCAGCTGCCGCCGGTG
 TTGGGAGGAGAAGTTGTATAACCCCCCCCACCCACCGACGGGTGAGCGGCGGGGCCACG

1681 AlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGly
 CGCTACTGCCATTGTGGCGCTGGCTTAGCTGGCGCCGCATCGGCAGTGTGGACTGG
 GCGATGACGGAAACACCCGCGACCGAATCGACCGCGGGTAGCCGTACAACCTGACCC

1741 LysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla
 GAAGGTCTCATAGACATCCTGCAAGGGTATGGCGGGCGTGGCGGGAGCTCTGTGGC
 CTTCCAGGAGTATCTGTTAGGAACGTCCCATTACCGCGCCCGCACCGCCCTCGAGAACACCG

1801 PheLysIleMetSerGlyGluValProSerThrGluAspLeuValAsnLeuLeuProAla
 ATTCAAGATCATGAGCGGTGAGGTCCCACGGAGGACCTGGTCAATCTACTGCCCGC
 TAAGTTCTAGTACTGCCACTCCAGGGAGGTGCCTGGACCAAGTTAGATGACGGGCG

1861 IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis
 CATCCTCTGCCGGAGCCCTCGTAGTCGGCGTGGCTGTGCAGCAATACTGCGCCGGCA
 GTAGGAGAGCGGGCCTGGGAGCATCAGCCGCACCAAGACACGTCGTTATGACGGGGCGT

FIG. 14-2 ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg

2212511

1921 CGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTCGCCTCCCG
 GCAACC GGCGCCGCTCCCCGTACGTACCTACTTGGCCGACTATCGGAAGCGGAGGGC
 GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr
 1981 GGGGAACCATGTTCCCCCACGCACACTACGTGCCGGAGAGCGATGCAGCTGCCCGCTCAC
 CCCCTTGGTACAAGGGGGTGCGTGATGCACGGCCTCTCGCTACGTCACGGCGCAGTG
 AlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSer
 2041 TGCCATACTCAGCAGCCTCACTGTAACCCAGCTCCTGAGGGCAGTGCACCACTGGATAAG
 ACGGTATGAGTCGTCGGAGTGACATTGGTCGAGGACTCCGCTGACGTGGTCACCTATTC
 SerGluCysThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCys
 2101 CTCGGAGTGTACCACTCCATGCTCCGGTCTGGCTAAGGGACATCTGGACTGGATATG
 GAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCCCTGACCTATAC
 GluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGly
 2161 CGAGGTGTTGAGCGACTTAAAGACCTGGCTAAAAGCTAACGCTCATGCCACAGCTGCCCTGG
 GCTCCACAACTCGCTGAAATTCTGGACCGATTTCGATTGAGTACGGTGTGACGGGACC
 IleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMet
 2221 GATCCCCTTGTCCTGCCAGCGCGGTATAAGGGGTCTGGCGAGTGGACGGCATCAT
 CTAGGGAAACACAGGACGGTCGCGCCATATTCCCCAGACCGCTACCTGCCGTAGTA
 HisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArg
 2281 GCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAACGGGACGATGAG
 CGTGTGAGCGACGGTACACCTCGACTCTAGTGACCTGTACAGTTTGCCCTGCTACTC
 IleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyr
 2341 GATCGTCGGTCTTAGGACCTGCAGGAACATGTGGAGTGGGACCTTCCCATTAAAGCCTA
 CTAGCAGCCAGGATCCTGGACGTCCATTGTACACCTCACCTGGAGGGTAATTACGGAT
 ThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgVal
 2401 CACCAACGGGCCCCCTGTACCCCCCTCTGCAGCGCAACTACAGTTCGCGCTAIGGAGGGT
 GTGGTCCCCGGGGACATGGGGGAAGGACGCGGCTGATGTGCAAGCGCGATACCTCCCA
 SerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMet
 2461 GTCTGCAGAGGAATATGTGGAGATAAGGCAGGTGGGGACTTCCACTACGTGACGGGTAT
 CAGACGTCTCTTACACCTCTATTCCGTCCACCCCTGAAGGTGATGCACGTCCATA
 ThrThrAspAsnLeuLysCysProCysGlnValProSerProGluPhePheThrGlu
 2521 GACTACTGACAATCTCAAATGCCGTGCCAGGTCCCCTCGCCGAATTTCACAGAAT
 CTGATGACTGTTAGAGTTACGGGACGGTCCAGGGTAGCGGGCTTAAAAAGTGTCTTA

FIG. 14-3

FIG. 15 Translation of DNA 33c

1 AlaValAspPheIleProValGluAsnLeuGluThrThrMetArgSerProValPheThr
 1 GGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAAACCATGAGGTCCCCGGTGTCAC
 CCGCCACCTGAAATAGGGACACCTCTGGATCTCTGGTACTCCAGGGGCCACAAGTG
 61 AspAsnSerSerProProValValProGlnSerPheGlnValAlaHisLeuHisAlaPro
 61 GGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCC
 CCTATTGAGGAGAGGTGGTCATCACGGGTCTCGAAGGTCCACCGAGTGGAGGTACGAGG
 121 ThrGlySerGlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLysVal
 121 CAACAGGCAAGCAGCGGCAAAGCACCAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGT
 GTGTCCCGTCGCCGTTTCTGGTCCAGGGCCGACGTACGTCGAGTCCCATAATCCA
 181 LeuValLeuAsnProSerValAlaAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAla
 181 GCTAGTACTCAACCCCTCTGGTCTGCAACACTGGGCTTGGTCTTACATGTCCAAGGC
 CGATCATGAGTTGGGAGACAACGACGTTGTGACCCGAAACACGAATGTACAGGTTCCG
 ----- Overlap with 40b -----
 241 HisGlyIleAspProAsnIleArgThrGlyValArgThrIleThrGlySerProIle
 241 TCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACAAATTACACTGGCAGCCCCAT
 AGTACCTAGCTAGGATTGTAGTCCTGGCCCCACTCTGGTAAATGGTGACCGTCGGGTA
 301 ThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAsp
 301 CACGTACTCCACCTACGGCAAGTTCTTGCCGACGGCGGGGTGCTCGGGGGCGTTATGA
 GTGCATGAGGTGGATGCEGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGCGAATACT
 361 IleIleIleCysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThr
 361 CATAATAATTGTGACGAGTGCCACTCCACGGATGCCACATCCATCTGGCATTGGCAC
 GTATTATTAAACACTGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAACCGTG
 421 ValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrPro
 421 TGTCTTGACCAAGCAGAGACTGGGGGGCGAGACTGGTGTGCTCGCCACCGCCACCC
 ACAGGAACCTGGTCTCTGACGCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGG
 481 ProGlySerValThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGly
 481 TCCGGGCTCCGTCACTGTGCCCCATCCAAACATCGAGGAGGTGCTCTGTCCACCACCG
 AGGCCCGAGGCAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCC
 541 GluIleProPheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeu
 541 AGAGATCCCTTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGACATCT
 TCTCTAGGGAAAATGCCATAGGGGGAGCTTCAATTAGTCCCCCTCTGTAGA
 601 IlePheCysHisSerLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGly
 601 CATCTCTGTCATTCAAAGAGAAAGTGCACGAACTCGCCGAAAGCTGGTGCATGGG
 GTAGAAAGACAGTAAGTTCTCTTACGCTGAGCGCGTTTCGACCGCGTAACCC
 661 IleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAsp
 661 CATCAATGCCGTGGCCTACTACCGCGGTCTTGACGTGTCATCCGACCAGCGGCCGA
 GTAGTTACGGCACCGGATGAGCGCCAGAACACTGACACAGGCAGTAGGGCTGGTCGCCGCT
 721 ValValValValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerVal
 721 TGTTGTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGGGCAGTCGACTCGGT
 ACAACAGCAGCACCGTTGGCTACGGGAGTACTGGCCGATATGGCGCTGAAGCTGAGCCA
 781 IleAspCysAsnThrCys
 781 GATAGACTGCAATACGTGTG
 CTATCTGACGTTATGCACAC

FIG. 16 Translation of DNA 8h

ProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIlePro
 1 CTCCCTGCACTTGGCGCTCCTCGGACCTTACCTGGTCACGAGGCACGCCGATGTCATTG
 GAGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGCGGCTACAGTAAG
 ValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeu
 61 CCGTGCGCCGGCGGGGTGATAGCAGGGGCAGCCTGCTGTCGCCCCGGCCATTCTACT
 GGCACGCCGCCCACTATCGTCCCCGTCGGACGACAGCGGGCCGGTAAAGGATGA
 LysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg
 121 TGAAAGGCTCCTCGGGGGTCCGCTGTTGCCCCGCGGGCACGCCGTGGCATATTAA
 ACTTTCCGAGGAGCCCCCAGGCACACACGGGCGCCCCGTGCGGCACCGTATAAAT

Overlap with

AlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIleProValGluAsnLeu
 181 GGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTATCCCTGTGGAGAAC
 CCCGGCGCACACGTGGCACCTACCGATTCCGCCACCTGAAATAGGGACACCTCTTGG

33c

GluThrThrMetArgSerProValPheThrAspAsnSer
 241 TAGAGACAACCATGAGGTCCCCGGTGTACGGATAACTCCTC
 ATCTCTGTGGTACTCCAGGGCCACAAGTGCCTATTGAGGAG

FIG. 17 Translation of DNA 7e

GlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGly
 1 GGGGTGGAGGTGCTGGCGCCCATCACGGCGTACGCCAGCAGACAAGGGCCTCTTAGG
 CCCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCCGTCTGTTCCCGAGGATCC

CysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIle
 61 GTGCATAATCACCAAGCCTAACTGGCGGGACAAAACCAAGTGGAGGGTGAGGTCCAGAT
 CACGTATTAGTGGTCGGATTGACCGGCCCTGTTTGGTACCTCCCCTCCAGGTCTA

ValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrVal
 121 TGTGTCAACTGCTGCCAACCTTCTGGCAACGTGCACTCAATGGGGTGCTGGACTGT
 ACACAGTTGACGACGGTTGGAAGGACCGTTGACGTAGTTACCCCACACGACCTGACA

TyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyr
 181 CTACCACGGGGCCGGAACGAGGACCATCGCGTACCCAAGGGCTGTCATCCAGATGTA
 GATGGTGCCCCGGCCCTGCTCTGGTAGCGCAGTGGTTCCCAGGACAGTAGGTCTACAT

ThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThr
 241 TACCAATGTAGACCAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTGAC
 ATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAACG

Overlap with 8h

ProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHis
 301 ACCCTGCACTTGGCGCTCCTCGGACCTTACCTGGTCACGAGGCACG
 TGGGACGTGAACGCCGAGGAGCCTGGAAATGGACCAGTGCTCCGTGC

FIG. 18 Translation of DNA 14c

AsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeu
 1 GPACATGTGGAGTGGGACCTCCCTTAATGCCTACACCACGGGCCCTGTACCCCCCT
 CTTGTACACCTCACCCCTGGAAGGGTAATTACGGATGTGGTGCCCAGGGACATGGGGGA

Overlap with 25c

ProAlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIle
 61 TCCTCGCCGAACTACACGTTGCGCTATGGAGGGTGTCTGCAGAGGAATACGTGGAGAT
 AGGACGCGCTTGATGTGCAAGCGCATACTCCCCACAGACGTCTCCTTATGCACCTCTA

ArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysPro
 121 AAGGCAGGTGGGGACTTCACTACGTGACGGGTATGACTACTGACAATCTAAATGCC
 TTCCGTCCACCCCCCTGAAGGTGATGCACTGCCACTACTGATGACTGTTAGAATTACGGG

CysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPhe
 181 GTGCCAGGTCCCCTCGCCCGAATTTTCACAGAATTGGACGGGGTGCCTACATAGGTT
 CACGGTCCAGGGTAGCGGGCTTAAAGTGTCTAACCTGCCACGCGATGTATCAA

AlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGlu
 241 TGC GCC CCC CTG CAAG CC CTG CTG CG GGG AGG AGGT ATC ATT CAG AGT TAGG ACT CC AC GA
 ACG CG GGG GGG AC GTT CG GG AA CG AC G CC CT CCT CC AT AG TA AGT CT CAT CCT GAGG TG CT

TyrProValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSer
 301 AT ACC CG GT AG GG TG CG CA ATT AC CT TG CG AG CC CG A AC CG GT GG CC GT TT GAC G TC
 TATGGGCCATCCCAGCGTTAATGGAACGCTGGGCTTGGCTGCACCGCACACTGCG

MetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGly
 361 CAT GCT CACT GAT CC CT CC CAT ATA AC AG CAG AGG C G G C G G C G A AG G T T G G C G A G G G G
 GTACGAGTGACTAGGGAGGGTATATTGCGTCTCGCCGGCCGTTCCAACCGCTCCCC

SerProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAla
 421 ATC ACC CCC CTG TG G CC AG C T C C G C T A G C C A G T A T C C G C T C C A T C T C A A G G C
 TAGTGGGGGGAGACACCGGTCGAGGAGCCGATCGGTGATAAGGCGAGGTAGAGAGTTCCG

ThrCysThrAlaAsnHisAspSerProAsp
 481 AACTTGCAACCGCTAACCATGACTCCCCCTGAT
 TTGAACGTGGCGATTGGTACTGAGGGGACTA

FIG. 19 Translation of DNA 8f

Overlap with 14c

1 SerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHis
 AGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACCGCTAACCAT
 TCGAGGAGCCGATCGTCGATAGGCGAGGTAGAGAGTTCCGTTAACGTGGCGATTGGTA

61 AspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGly
 GACTCCCCTGATGCTGAGCTATAGAGGCCAACCTCCTATGGAGGCAGGAGATGGGCGGC
 CTGAGGGGACTACGACTCGAGTATCTCCGGTGGAGGATACCTCCGTCTACCCGCCG

121 AsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSerPheAspProLeu
 AACATCACCAAGGGTTGAGTCAGAAAACAAGTGGTATTCTGGACTCCTCGATCCGCTT
 TTGTAGTGGTCCCAACTCAGTCTTGTGTTACCACTAAGACCTGAGGAAGCTAGGCGAA

181 ValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeuArgLysSerArg
 GTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGAGAAATCCTGCGGAAGTCTCGG
 CACCGCCTCCTCGTCCGCCCCTCTAGAGGCATGGCGTCTTAGGACGCCCTCAGAGCC

241 ArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnProProLeuValGlu
 AGATTGCCAGGCCCTGCCGTTGGCGCGCCGGACTATAACCCCCCGTAGTGGAG
 TCTAACGGGTCCGGGACGGGCAAACCCGCGCCGGCTGATATGGGGGCGATCACCTC

301 ThrTrpLysProAspTyrGluProProValValHisGlyCysProLeuProProPro
 ACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTTCCACCTCCA
 TGACACCTTTCTGGCTGATGCTTGGACACCAGGTACCGACAGGGGAAGGTGGAGGT

361 LysSerProProValPro
 AAGTCCCCCTCGTGC
 TTCAGGGGAGGACACGGC

FIG. 20 Translation of DNA 33f

1 ValTrpAlaArgProAspTyrAsnProProLeuValGluThrTrpLysLysProAspTyr
 CGTTGGCGCGCCGGACTATAACCCCCCGTAGTGGAGACGTGGAAAAAACCGACTA
 GCAAACCCGCGCCGGCTGATATGGGGGCGATCACCTCTGACCTTTTGGCTGAT

Overlap with 8f

61 GluProProValValHisGlyCysProLeuProProProLysSerProProValProPro
 CGAACACCTGTGGTCCATGGCTGCCGCTTCCACCTCCAAAGTCCCCTCTGTGCTCC
 GCTTGGTGGACACCAGGTACCGACGGCGAAGGTGGAGGTTCAAGGGAGGACACGGAGG

121 ProArgLysLysArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAlaGlu
 GCCTCGGAAGAACGGACGGTGGTCCTCACTGAATCAACCTATCTACTGCTTGGCCGA
 CGGAGCCTCTCGCCTGCCACCAGGAGTGAATTAGTTGGATAGATGACGGAACCGGCT

181 LeuAlaThrArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThr
 GCTCGCCACCAAGCTTGGCAGCTCCTCAACTTCCGGCATTACGGGCGACAATACGAC
 CGAGCGGTGGTCTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCGCTGTTATGCTG

241 ThrSerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerPhe
 AACATCCTCTGAGCCGCCCTCTGGCTGCCCCCCCAGCTCCGACGCTGAGTCCTTGC
 TTGTAGGAGACTCGGGCGGGGAAGACCGACGGGGGGCTGAGGCTGCGACTCAGGAAACG

FIG. 21 Translation of DNA 33g

1 AlaSerArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThrThr
 GCCTCCAGAAGCTTGGCAGCTCTCAACTCCGCATTACGGCGACAATACGACAACA
 CGGAGGTCTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCGCTGTTATGCTGTTGT

Overlap with 33f-----

61 SerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSer
 TCCTCTGAGCCCCCCCCTCTGGCTGCCCCCGACTCCGACGCTGAGTCCTATTCCCTCC
 AGGAGACTCGGGGGGGAAAGACCGACGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGG

121 MetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThr
 ATGCCCCCCCCTGGAGGGGGAGCCTGGGATCCGGATCTTAGCGACGGGTATGGTCACAG
 TACGGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCAGTACCAAGTTGC

181 ValSerSerGluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThr
 GTCAGTAGTGAGGCCAACCGGGAGGATCGTGTGCTCAATGTCTTACTCTGGACA
 CAGTCATCACTCCGGTTCGCCTCCTACAGCACACGACGAGTTACAGAATGAGAACCTGT

241 GlyAlaLeuValThrProCysAlaAlaGluGlnLysLeuProIleAsnAlaLeuSer
 GGCGCACTCGTCACCCCGTGCACGGCGGAAGAACAGAAACTGCCCCTCAATGCACATAAGC
 CCGCGTGAGCAGTGGGGCACCGCGCCCTCTGTCTTGACGGGTAGTTACGTGATTG

301 AsnSerLeuLeuArgHisHisAsnLeuValTyrSerThrThrSerArgSer
 AACTCGTTGCTACGTCACCACAATTGGTGTATTCCACCACCTCACGCAGTG
 TTGAGCAACGATGCAGTGGTAAACACACATAAGGTGGTAGTGCAC

FIG. 22 Translation of DNA 7f

1 GlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArg
 GGCACCTATGTTATAACCATCTCACCTCTCGGACTGGCGCACACGGCTGCGA
 CCGTGATACAAATATTGGTAGAGTGAGGAGAACCCCTGACCGCGTGGTGCACGCT

61 AspLeuAlaValAlaValGluProValValPheSerGlnMetGluThrLysLeuIleThr
 GATCTGGCCGTGGCTGTAGAGCCAGTCGTCCTCTCCCAAATGGAGACCAAGCTCATCAGC
 CTAGACCGGCACCGACATCTCGGTAGCAGAACAGAGGGTTACCTCTGGTGCAGTAGTGC

121 TrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArg
 TGGGGGGCAGATACCGCCGCGTGCAGGTGACATCATCAACGGCTGCGTGGTCCGCCGC
 ACCCCCCGTCTATGGCGCGCACGCCACTGTAGTAGTTGCCGAACGGACAAAGGCCGGCG

181 ArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeu
 AGGGGCCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCCAAGGGTTGGAGGTTG
 TCCCCGGCCCTCTATGACGAGCCGGTCGGTACCTTACCAAGAGGTTCCAACCTCCAAC

241 LeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr
 CTGGCGCCCATCACGGCGTACGCCAGCAGACAAGGGGCCTCTAGGGTGCATAATCACC
 GACCGCGGGTAGTGCCTGCGCATGCGGTGCGTCTGTTCCCCGGAGGATCCCACGTATTAGTGG

Overlap with 7e-----

301 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla
 AGCCTAACTGGCCGGACAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT
 TCGGATTGACCGCCCTGTTGGTACCTCCACTCCAGGTCTAACACAGTTGACGA

361 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrp
 GCCCAAACCTTCTGGCAACGTGCATCAATGGGTGTGCTGG
 CGGGTTGGAGGACCGTTGCACGTAGTTACCCACACGACC

FIG. 23 Translation of DNA 11b

GlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyr
 1 GGC GGTTGTTCTCGTCGGGTTGATGGCGCTGACTCTGTCACCATATTACAAGCGCTAT
 CCGCCACAAACAAGAGCAGCCCAACTACCGCGACTGAGACAGTGGTATAATGTTCGCGATA

 IleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHis
 61 ATCAGCTGGTGCTTGTGGCTTCAGTATTTCTGACCAGAGTGGAAGCGCAACTGCAC
 TAGTCGACCAACGAAACACCACCGAAGTCATAAAAGACTGGTCTCACCTCGCGTTGACGTG

 ValTrpIleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeuMetCys
 121 GTGTGGATTCCCCCCCCTCAACGTCCGAGGGGGCGCGACGCCGTATCTTACTCATGT
 CACACCTAACGGGGGAGTTGCAGGCTCCCCCGCGCTGCCGAGTAGAATGAGTACACA

 AlaValHisProThrLeuValPheAspIleThrLysLeuLeuLeuAlaValPheGlyPro
 181 GCTGTACACCCGACTCTGGTATTGACATCACCAAATTGCTGCTGCCGTCTCGGACCC
 CGACATGTGGGCTGAGACCATAAACTGTAGTGGTTAACGACGACCGGCAGAACGCCTGGG

 LeuTrpIleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGlnGlyLeu
 241 CTTGGATTCTTCAAGCCAGTTGCTAAAGTACCCACTTTGTGCGCGTCCAAGGCCTT
 GAAACCTAACGAAAGTTCGGTAAACGAATTTCATGGGATGAAAACACGCGCAGGTTCCGGAA

 LeuArgPheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMetValIle
 301 CTCCGGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATC
 GAGGCCAAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCACGTTACCACTAG

 IleLysLeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeuArgAsp
 361 ATTAAGTTAGGGCGCTACTGGCACCTATGTTATAACCATCTCACTCCTCTCGGGAC
 TAATTCAATCCCCCGCAATGACCGTGGATACAAATATTGGTAGAGTGAGGAGAACGCCCTG

 -----Overlap with 7f-----
 TrpAlaHisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPheSerGln
 421 TGGGCGCACACGGCTTGCAGAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCAA
 ACCCGCGTGTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTAGCAGCAGAACAGGGTT

 MetGluThrLysLeuIleThrTrpGly
 481 ATGGAGACCAAGCTCATCACGTGGGGGC
 TACCTCTGGTTGAGTAGTGCACCCCCCG

FIG. 24 Translation of DNA 14i

GluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrp
 1 GGGAGTACGTCGTTCTCCTGTTCTGCTGAGACGCGCGGTCTGCTCCTGCTTG
 CCCTCATGCAGCAAGAGGAAGACGAACGCTCTGCGCGCAGACGAGGACAAACA
 MetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAla
 61 GGATGATGCTACTCATATCCCAAGCGGAGGCGGTTGGAGAACCTCGTAATACTTAATG
 CCTACTACGATGAGTATAGGGTTCGCCTCCGCGAACACCTCTGGAGCATTATGAATTAC
 AlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrp
 121 CAGCATCCCTGGCCGGGACGACGGTCTTGATCCCTCGTGTCTGCTTTGCAT
 GTCGTAGGGACCGGCCCTCGGTGCCAGAACATAGGAAGGAGCACAAAGAACGAAACGTA
 TyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeu
 181 GGTATTGAAGGTAAGTGGGTGCCCGAGCGGTCTACACCTCTACGGGATGTGGCTC
 CCATAAACTTCCCATTCAACCCACGGCCTGCCAGATGTGAAAGATGCCCTACACCGGAG
 LeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAla
 241 TCCTCCCTGCTCCTGTTGGCGTTGCCCGAGCGGGCGTACGCGCTGGACACGGAGGTGGCCG
 AGGAGGACGAGGACAACCGCAACGGGTCGCCGATGCGCACCTGTGCCCTCCACCGGC
 -----Overlap with 1lb-----
 SerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLys
 301 CGTCGTGTGGCGGTGTTGTTCTCGTGGGTGACTGCGCTGACTCTGTCACCATATTACA
 GCAGCACACCAGCACAAGAGCAGCCAACCTACCGCAGTGGACAGTGGTATAATGT

 ArgTyrIleSerTrpCysLeuTrpTrpLeuGln
 361 AGCGCTATATCAGCTGGTGTGTTGGCTCAGAA
 TCGCGATATAGTCGACCACGAACACCACCGAAGTCTT

FIG. 25 Translation of DNA 39c

ProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProPro
 1 CCAGCCCCCTGGCTGCCCGACTCCGACGCTGAGTCTTCTCCATGCCCG
 GGTCGGGAAGACCGACGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGG

 LeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSer
 61 CTGGAGGGGGAGCCTGGGATCCGATCTTAGCGACGGGTCTGGTCAACAGTCAGTAGT
 GACCTCCCCCTCGGACCCCTAGGCCTAGAACCGCTGCCCAGTACCGTTGTCAGTCATCA
 -----Overlap with 33g-----
 GluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeu
 121 GAGGCCAACCGCGAGGATGTCGTGCTGCTCAATGTCCTACTCTGGACAGGCGCAGTC
 CTCCGGTTGCGCCTCCTACAGCACACGAGTTACAGGATGAGAACCTGTCAGTCATCA

 ValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeu
 181 GTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCATCAATGCACTGAGCAACTCGTG
 CAGTGGGGCACCGCGCCTTCTGTCTTGACGGTAGTTACGTGACTCGTTGAGCAAC

 LeuArgHisAsnLeuValTyrSerThrSerArgSerAlaCysGlnArgGlnLys
 241 CTACGTACCCACAATTGGGTATTCCACACCTCACGAGTGCTTGCAAAGGCAGAAG
 GATGCACTGGTGTAAACACATAAGGTGGTAGTGCACGAGTACGGTTCCGTCTTC

 LysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGlu
 301 AAAGTCACATTGACAGACTGCAAGTTCTGGACAGCATTACAGGACGTACTCAAGGAG
 TTTCAGTGTAAACTGTCGTTCAAGACCTGCGTAATGGTCCCTGCATGAGTTCCCTC

 ValLysAlaAlaAlaSerLysValLysAlaAsnPhe
 361 GTTAAAGCAGCGCGTCAAAAGTGAAGGCTAACTTC
 CAATTTCGTCGCCAGTTTCACTTCCGATTGAAG

FIG. 26-1 COMBINED ORF OF DNAs
14i/11b/7f/7e/8h/33c/40b/37b/35/36/81/32/33b/25c/14c/8f/33f/33g/39c

1 GluTyrValValLeuLeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrp
 1 GGGAGTACGTCGTTCTCTGTTCCCTCTGCTTGAGACGCGCGCTGCTCCCTGCTTGT
 1 CCCTCATGCAGCAAGAGGAAGACGAACGTCTGCGCGQGCAGACGAGGACGAACA

 61 MetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAla
 61 GGATGATGCTACTCATATCCCAAGCGGAGGCCGTTGGAGAACCTCGTAATACTTAATG
 61 CCTACTACGATGAGTATAGGTTGCCCTCGCCGAAACCTCTGGAGCATTATGAATTAC

 121 AlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrp
 121 CAGCATCCCTGGCCGGACGCACGGCTTGATCCTCCTCGTGTCTGCTTGCAT
 121 GTCGTAGGGACCAGGCCCTCGCTGCCAGAACATAGGAAGGAGCACAAAGAACGAA

 181 TyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeu
 181 GGTATTTGAAAGGCTAAGTGGGTGCCGGAGCGGTCTACACCTCTACGGGATGTGGCCTC
 181 CCATAAACTTCCCATTCAACCCACGGCCTGCCAGATGTGGAAGATGCCCTACACCGGAG

 241 LeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAla
 241 TCCCTCCTGCTCTGTTGGCGTTGCCAGCGGGCTACGGCCTGGACACGGAGGTGGCCG
 241 AGGAGGACGAGGACAACCGCAACGGGTCGCCGCATGCCGACCTGTGCCCTCACCGGC

 301 SerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLys
 301 CGTCGTGCGGGTGTGTTCTGCGGGTTGATGGCGCTGACTCTGTCACCATATTACA
 301 GCAGCACACCGCCACAACAGAGCAGCCAACTACCGCGACTGAGACAGTGGTATAATGT

 361 ArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGln
 361 AGCGCTATATCAGCTGGTGCTTGTTGGCTTCAGTATTTCTGACCATGAGTGGAAAGCGC
 361 TCGCGATATAGTCGACCAACGAAACACCACCGAAGTCATAAAAGACTGGTCTCACCTCGCG

 421 LeuHisValTrpIleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeu
 421 AACTGCACGTGTGGATTCCCCCCCCTCAACGTCGAGGGGGCGCAGCCGTATCTAC
 421 TTGACGTGACACCTAACGGGGGGAGTTGCAGGCTCCCCCGCGTGCAGTAGAATG

 481 MetCysAlaValHisProThrLeuValPheAspIleThrLysLeuLeuLeuAlaValPhe
 481 TCATGTGTGCTGTACACCGACTCTGGTATTGACATCACCAAATTGCTGCTGCCGTCT
 481 AGTACACACGACATGTGGCTGAGACCATAAACACTGTAGTGGTTAACGACGACCGGCAGA

 541 GlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGln
 541 TCGGACCCCTTGGATTCTCAAGCCAGTTGCTTAAAGTACCCACTTTGTGCCGTCTC
 541 AGCCTGGGAAACCTAACAGTTGGCTAACGAATTTCATGGGATGAAACACCGCGCAGG

 601 GlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMet
 601 AAGGCCTTCCTCGGTTCTGCGCGTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAAA
 601 TTCCGGAAGAGGCCAACGCGCAATCGCCCTCTACTAGCCTCCGGTAATGCACGTT

 661 ValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeu
 661 TGTCATCATTAAGTTAGGGCGCTTACTGGCACCTATGTTATAACCATCTCACTCCTC
 661 ACCAGTAGTAATTCAATCCCCCGCAATGACCGTGGATAACAAATTGGTAGAGTGAGGAG

 721 ArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPhe
 721 TTGGGGACTGGGCGCACACGGCTTGCAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCT
 721 AACCCCTGACCCCGCGTGTGCCGAACGCTCTAGACCGGACCGACATCTGGTCAGCAGA

 781 SerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGlyAspIle
 781 TCTCCCAAATGGAGACCAAGCTCATCACGTGGGGGGCAGATACCGCCCGTGGGTGACA
 781 AGAGGGTTTACCTCTGGTCAGTAGTCGACCCCCCGTCTATGGCGCGCACGCCACTGT

 841 IleAsnGlyLeuProValSerAlaArgArgGlyArgGluIleLeuGlyProAlaAsp
 841 TCATCAACGGCTTGCCTGTTCCGCCCGCAGGGGGCAGGAGATACTGCTCGGGCCAGCCG
 841 AGTAGTTGCCGAACGGACAAAGCGGGCGTCCCCGGCCCTATGACGAGCCCCGGTCGGC

GlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThr

901 ATGGAATGGTCTCCAAGGGGTGGAGGTTGCTGGCGCCCATCACGGCGTACGCCAGCAGA
 TACCTTACCAAGAGGTTCCCCACCTCAACGACCGCGGTAGTGCCGCATGCCGGTCGTCT
 ArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGlu
 961 CAAGGGGCCTCCTAGGGTGCATAATACCAGCCTAACTGGCCGGACAAAAACCAAGTGG
 GTTCCCCGGAGGATCCCACGTATTAGTGGTCGGATTGACCGGCCCTGTTGGTTCAAC
 GlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGly
 1021 AGGGTGTAGGTCCAGATTGTGTCACTGCTGCCAACCTTCTGGCAACGTGCATCAATG
 TCCCACCTCCAGGTCTAACACAGTTGACGGACGGGTTGGAAGGACCGTGCACGTAGTTAC
 ValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyPro
 1081 GGGTGTGCTGGACTGTCTACACACGGGGCCGGAACGAGGACCATCGCGTACCCAAAGGGTC
 CCCACACGACACTGACAGATGGTGCCTGGCTGGTAGCGCAGTGGGTTCCAG
 ValIleGlnMetTyrThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGly
 1141 CTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGGCCGCTCCGCAAG
 GACAGTAGGTCTACATATGGTTACATCTGGTCTGGAACACCCGACCGGGCGAGGCGTTC
 SerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHis
 1201 GTAGCCGCTCATGACACCCCTGCACCTGCGGCTCTCGGACCTTACCTGGTCACGAGGC
 CATCGGCGAGTAACGTGGGACGTGAACGCCAGGAGCCTGAAATGGACAGTGCCTCCG
 AlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArg
 1261 ACGCCGATGTCATTCCCGTGCACGGCGGGGTGATAGCAGGGCAGCCTGCTGCGCCCG
 TGCGGCTACAGTAAGGGCACCGCGGCCACTATCGTCCCCGTCGGACGACAGCGGGG
 ProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAla
 1321 GGCCCATTCTACTTGAAAGGCTCTCGGGGGTCCGCTGTGTGCCCCGGGGCACG
 CGGGTAAAGGATGAACCTTCCGAGGAGCCCCCAGGCGACAACACGGGGCGCCCCGTGC
 ValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIle
 1381 CCGTGGGCATATTAGGGCCGGTGTGCACCGTGGAGTGGCTAAGGGGGTGGACTTTA
 GGCACCCGTATAATCCGGCGCACACGTGGGCACCTCACCGATTCCGCCACCTGAAAT
 ProValGluAsnLeuGluThrThrMetArgSerProValPheThrAspAsnSerSerPro
 1441 TCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTACGGATAACTCCCTCTC
 AGGGACACCTCTGGATCTGTGGTACTCCAGGGCCACAAGTGCCTATTGAGGAGAG
 ProValValProGlnSerPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLys
 1501 CACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATGCTCCCACAGGCAGCGGCA
 GTGGTCATCACGGGTCTGAAGGTCCACCGAGTGGAGGTACGGGTGTCCGTCGGCGT
 SerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsnPro
 1561 AAAGCACCAAGGTCCCCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACC
 TTCTGTGGTCCAGGGCGACGTATACGTCGAGTCCCAGATATTCCACGATCATGAGTTGG
 SerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAspPro
 1621 CCTCTGTGCTGCAACACTGGGTTTGCTTACATGTCCAAGGCTCATGGATCGATC
 GGAGACAACGACGTTGTGACCCGAAACACGAATGTACAGGTTCCGAGTACCCCTAGCTAG
 AsnIleArgThrGlyValArgThrIleThrGlySerProIleThrTyrSerThrTyr
 1681 CTAACATCAGGACCGGGGTGAGAACAAATTACCACTGGCAGCCCCATCACGTACTCCACCT
 GATTGTAGTCCTGGCCCCACTCTGTTAATGGTGACCGTCGGGTAGTGCATGAGGTGG
 GlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCysAsp
 1741 ACGGCAGTTCTTGCCGACGGCGGGTGTGCTGGGGCGCTTATGACATAATAATTGTG
 TGCCGTTCAAGGAACGGCTGCCGCCACGAGCCCCCGCGAATACTGTATTATAAACAC
 GluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGlnAla
 1801 ACAGAGTGCACCGGATGCCACATCCATCTGGCATGGCACTGTCCCTGACCAAG
 TGCTCACGGTGAGGTGCCTACGGTGTAGGTAGAACCCGTAGCCGTGACAGGAACGGTTC
 GluThrAlaGlyAlaArgLeuValLeuAlaThrAlaThrProProGlySerValThr
 1861 CAGAGACTGCCGGGGCGAGACTGGTTGTGCTGCCACCGCCACCCCTCCGGCTCCGTCA
 GTCTCTGACGCCCGCTGACCAACACGAGCGGTGGCGTGGGGAGGCCGAGGCAGT

1921 ValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPheTyr
 CTGTCCCCATCCCAACATCGAGGAGGTTGCTCTGTCACCACCGGAGAGATCCCTTTT
 GACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTAGGGAAAAA

 1981 GlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHisSer
 ACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTGTGTCATT
 TGGCGTTCCGATAGGGGAGCTTCATTAGTCCCCCTCTGTAGAGTAGAACAGATAA

 2041 LysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaValAla
 CAAAGAAGAAGTGCAGACTCGCCGAAAGCTGGTCGATTGGGCATCAATGCCGTGG
 GTTCTTCTCACGCTGCTTGAGCGCGTTCGACCAGCGTAACCGTAGTTACGGCACC

 2101 TyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValValAla
 CCTACTACCGCGGTCTTGACGTGTCGTCATCCGACCAGCGCGATGTTGTCGTCGTGG
 GGATGATGGCGCCAGACTGCACAGGCACTAGGGCTGGTCGCCGCTAACACAGCAGCACC

 2161 ThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsnThr
 CAACCGATGCCCTCATGACCGGCTATACCGCGACTTCGACTCGGTAGACTGCAATA
 GTTGGCTACGGGAGTACTGGCGATATGGCGCTGAAGCTGAGCAGTACTGACGTTAT

 2221 CysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThr
 CGTGTGTCACCCAGACAGTCGATTTCAGCCTGACCCCTACCTCACCATIGAGACAATCA
 GCACACAGTGGTCTGTCAGCTAAAGTCGAACTGGGATGGAAGTGGTAACCTGTTAGT

 2281 LeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLysPro
 CGCTCCCCCAGGATGCTGTCGGCACTCAACGTGGGGCAGGACTGGCAGGGGGAAAGC
 GCGAGGGGTCCTACGACAGAGGGCGTGAGTGCAGCCCCGTCCTGACCGTCCCCCTCG

 2341 GlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSerVal
 CAGGCATCTACAGATTGTGGCACCGGGGAGCGCCCTCCGGCATGTTGACTCGTCCG
 GTCCGTAGATGTCTAACACCGTGGCCCCCTCGCGGGGAGGCCGACAAGCTGAGCAGGC

 2401 LeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThrThr
 TCCTCTGTGAGTGCTATGACGCAGGCTGCTTGATGAGCTACGCCGCGAGACTA
 AGGAGACACTACGATACTGCGTCCGACACGAACCATACTCGAGTGCAGGGCTCTGAT

 2461 ValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGlu
 CAGTTAGGCTACGAGCTACATGAAACACCCGGGGCTTCCCGTGTGCCAGGACCATCTTG
 GTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAAGGGCACACGGTCTGGTAGAAC

 2521 PheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThr
 AATTTGGGAGGGCGTCTTACAGGCCTCACTCATATAGATGCCACTTCTATCCCAGA
 TTAAAACCTCCCGCAGAAATGCCGGAGTGAGTATCTACGGGTGAAAGATAGGGTCT

 2581 LysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAlaArg
 CAAAGCAGGTGGGAGAACCTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTA
 GTTCTGTCACCCCTCTGGAAAGGAATGGACCATCGCATGGTTCGGTGGCACACGCGAT

 2641 AlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLysPro
 GGGCTCAAGCCCCCTCCCCCATCGTGGGACCAGATGTTGATTGCGCTCAAGC
 CCCGAGTTGGGGAGGGGGTAGCACCCTGGTCTACACCTTACAAACTAACGGAGTTG

 2701 ThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIleThr
 CCACCCCATGGGCCAACACCCCTGCTATACAGACTGGCGCTGTTGAGAATGAAATCA
 GGTGGGAGGTACCCGGTGTGGGGACGATATGTCGACCCCGCACAAGTCTTACTTGT

 2761 LeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluValVal
 CCCTGACGCACCCAGTCACCAAATACATGACATGTCGGCCGACCTGGAGGTG
 GGGACTGCGTGGGTCACTGGTTATGAGTACTGTACGTCAGCCGGCTGGACCTCCAGC

 2821 ThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSer
 TCACGAGCACCTGGGTGCTCGTTGGCGGGCTCTGGCTGTTGGCCGCGTATTGCGCTG
 AGTGCTGTTGGACCCACGAGCAACCGCCGAGGACCGACGAAACCGGGCGATAACGGACA
 AGTGCTGTTGGACCCACGAGCAACCGCCGAGGACCGACGAAACCGGGCGATAACGGACA
 AGTGCTGTTGGACCCACGAGCAACCGCCGAGGACCGACGAAACCGGGCGATAACGGACA

2881 CAACAGGCTCGTGGTCATAGTGGCAGGGTCGTCTTGTCCGGAAAGCCGGCAATCATAC
 GTTGTCCGACGCACCAGTATCACCCGTCCCAGCAGAACAGGCCCTCGGCCGTTAGTATG
 AspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuPro
 2941 CTGACAGGGAAAGTCTCTACCGAGAGTTCGATGAGATGGAAGAGTGTCTCAGCACTTAC
 GACTGTCCCTCAGGAGATGGCTCTCAAGCTACTCTACCTCTCACGAGAGTCGTGAATG
 TyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeuLeu
 3001 CGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAACGGCCCTCGGCCCTCC
 GCATGTAGCTCGTCCCTACTACGAGCGCTCGTCAAGTGTCTCCGGAGGCCGGAGG
 GlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGlnLys
 3061 TGCAGACCGCGTCCCGTCAGGCAGAGGTATCGCCCCGTGTCAGCAGAACCTGGCAA
 ACGTCTGGCGCAGGGCAGTCCGTCCAATAGCGGGGACAGGGTCTGGTTGACCGTTT
 LeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeuAla
 3121 AACTCGAGACCTTCTGGCGAAGCATAATGGAACATTCACTAGTGGATACAATACTTGG
 TTGAGCTCTGGAAGACCCGTTCTGTATACACCTGAAGTAGTCACCTATGTTATGAACC
 GlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAlaAla
 3181 CGGGCTGTCAACGCTGCCGGTAACCCGCCATTGCTTCATTGATGGCTTTACAGCTG
 GCCCGAACAGTGTGCGACGGACCATTGGGGCGGTAAACGAAGTAACCGAAAATGTCGAC
 ValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrpVal
 3241 CTGTCACCAGCCCCACTAACCACTAGCCAAACCCCTCTTCAACATATTGGGGGGGGGG
 GACAGTGGTCGGGTGATTGGTGATCGGTTGGGAGGAGAAGTTGTATAACCCCCCCCACCC
 AlaAlaGlnLeuAlaAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAlaGly
 3301 TGGCTGCCAGCTGCCGCCGGTGCCTACTGCCCTTGTTGTCAGGCTGGCTAGCTG
 ACCGACGGGTCAGCGGGCGGGGCCACGGCGATGACGGAAACACCCGCGACCGAATCGAC
 AlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyrGly
 3361 GCGCCGCCATCGGCAGTGTGGACTGGGAAGGTCTCATAGACATCCTGCAGGGTATG
 CGCGGGCGTAGCGTCACAACCTGACCCCTCCAGGAGTATCTGTAGGAACGTCCCATAAC
 AlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSerThr
 3421 GCGCGGGCGTGGCGGGAGCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCCCTCCA
 CGCGCCCCGACCGCCCTCGAGAACACCGTAAGTTCTAGTACTGCCACTCAGGGGAGGT
 GluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGlyVal
 3481 CGGAGGACCTGGTCAATCTACTGCCGCCATCCTCTGCCCGGAGCCCTCGTAGTCGGCG
 GCCTCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGGCTCGGGAGCATCAGCCGC
 ValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrpMet
 3541 TGGCTGTGCAGCAATACTGCCGGCACGTTGGCCGGCGAGGGGGCAGTGCAGTGG
 ACCAGACACGTCGTTATGACGGCGGTGCAACCGGGCCGCTCCCCGTACGTCACCT
 AsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrValPro
 3601 TGAACCGGCTGATAGCCTCGCCTCCGGGGAACCATGTTCCCCCACGCACTACGTGC
 ACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGGGTGCCTGATGCACG
 GluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGlnLeu
 3661 CGGAGAGCGATGCGAGCTGCCCGTCACTGCCATACTCAGCAGCCTACTGTAACCCAGC
 GCCTCTCGCTACGTCGACGGCGCAGTGAACGGTATGAGTCGTCGGAGTGACATTGGTCG
 LeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGlySerTrp
 3721 TCCTGAGGCGACTGCACCAGTGGATAAGCTCGGAGTGTACCACTCCATGCTCCGGTTCC
 AGGACTCCGCTACGTCGAGTGGTACCTATTGAGCCTCACATGGTGAGGTACGAGGCCAAGGA
 LeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeuLys
 3781 GGCTAAGGGACATCTGGACTGGATATGCGAGGTGTTGAGCGACTTTAACGACTGGCTAA
 CCGATTCCCTGTAGACCCCTGACCTATACGCTCCACAACTCGCTGAAATTCTGGACCGATT
 AlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyrLys
 3841 AAGCTAACGCTCATGCCACAGCTGCCGGATCCCCTTGTGTCTGCCAGCGGGTATA
 TTCGATTGAGTACGGTGTGACGGACCCCTAGGGGAAACACAGGACGGTCGCGCCCATAT

FIG. 26-4

2212511

26

GlyValTrpArgValAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIleThr
3901 AGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCA
TCCCCCAGACCGCTCACCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAGT

GlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMetTrp
3961 CTGGACATGTCAAAACGGGACGATGAGGATCGTCGGTCTAGGACCTGCAGGAACATGT
GACCTGTACAGTTTGCCCTGCTACTCCTAGCAGCCAGGATCTGGACGTCCCTGTACA

SerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAlaPro
4021 GGAGTGGGACCTTCCCCATTATGCCCTACACCACGGGCCCTGTACCCCCCTCCTGC
CCTCACCTGGAAGGGTAATTACGGATGTGGTGCCGGGACATGGGGGAAGGACGCG

AsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGlnVal
4081 CGAACTACACGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATAATGTGGAGATAAGGCAGG
GCTTGATGTGCAAGCGCGATAACCTCCCACAGACGTCTCCTATACACCTTATTCCGTCC

GlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGlnVal
4141 TGGGGGACTTCAACTACGTGACGGGTATGACTACTGACAATCTAAATGCCGTGCCAGG
ACCCCCTGAAGGTGATGCACTGCCCATACTGATGACTGTTAGAGTTACGGGCACGGTCC

ProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaProPro
4201 TCCCCATCGCCCCGAAATTTCACAGAATTGGACGGGGTGCCTACATAGGTTGCGCC
AGGGTAGCGGGCTTAAAAGTGTCTAACCTGCCACGCGGATGTATCAAACCGCGGG

CysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyrProVal
4261 CCTGCAAGCCCTGCTGCGGGAGGAGGTATTCAGAGTAGGACTCCACGAATACCCG
GGACGTTGGGAACGACGCCCTCCATAGTAAGTCTCATCCTGAGGTGCTTATGGGCC

GlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMetLeuThr
4321 TAGGGTGCCTAACCTTGCAGGCCCCAACCGGACGTGGCGTTGACGTCCATGCTCA
ATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGACAACGTGAGGTACGAGT

AspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerProPro
4381 CTGATCCCTCCCATAAACAGCAGAGGCCGGCGGGCGAACGGGATCACCC
GACTAGGGAGGGTATATIGTCGTCTCCGCCGGCCGCTTCAACCGCTCCCTAGTGGGG

SerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCysThr
4441 CCTCTGTGGCCAGCTCCTCGGTAGCCAGTATCCGCTCCATCTCAAGCAACTGCA
GGAGACACCGGTCGAGGAGCCATCGGTGAGGTAGAGAGTTCCGTTAACGT

AlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGlnGlu
4501 CCGCTAACCATGACTCCCCGTAGCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGG
GGCATTGGTACTGAGGGACTACGACTCGAGTATCTCCGGTTGGAGGATACCTCCGTCC

MetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAspSerPhe
4561 AGATGGCGGCAACATCACCAAGGGTTGAGTCAGAAAACAAAGTGGTATTCTGGACTCCT
TCTACCCGCCGTGTAGTGGTCCAACTCAGTCTTTGTTTACCAACTAACGACTGAGGA

AspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIleLeuArg
4621 TCGATCCGCTGTGGCGGAGGAGGACGAGCGGGAGATCTCGTACCCGCAAAATCCTGC
AGCTAGGCAGAACACCGCCTCCCTGCTGCCCTAGAGGCATGGCGTCTTAGGACG

LysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnProPro
4681 GGAAGTCTCGGAGATTGCCCAAGGCCCTGCCGTGGCGCGGACTATAACCCCC
CCTTCAGAGCCTCTAACGGGGTCCGGGACGGGAAACCCGCGCCGCGCTGATATTGGGG

LeuValGluThrTrpLysProAspTyrGluProProValValHisGlyCysProLeu
4741 CGCTAGTGGAGACGTGGAAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTC
GCGATCACCTCTGCACCTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACAGGCG

ProProProLysSerProProValProProProArgLysArgThrValValLeuThr
4801 TTCCACCTCCAAAGTCCCCCTCTGTGCCCTCGGCCCTCGGAAGAACGGGACGGTGGTCTCA
AAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTCTCGCCCTGCCACCAAGGAGT

GluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSerSer

FIG. 26-5

C

27

2212511

4861 CTGAATCAACCCTATCTACTGCCTGGCCGAGCTGCCACCAGAACGCTTGGCAGCTCCT
GACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGCTTCGAAACCGTCGAGGA
ThrSerGlyIleThrGlyAspAsnThrThrSerSerGluProAlaProSerGlyCys
4921 CAACTTCCGGCATACGGGCGACAATACGACAACATCCTCTGAGCCCCCCCCTCTGGCT
GTTGAAGGCCGTAATGCCGCTGTTATGCTGTTAGGAGACTCGGGCGGGGAAGACCGA
ProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGluProGly
4981 GCCCCCCCGACTCCGACGCTGAGTCCTATTCCCTCATGCCCGCCCTGGAGGGGGAGCCTG
CGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTCGGAC
AspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGluAsp
5041 GGGATCCGGATCTAGCGACGGGTATGGTCAACGTCAGTAGTGAGGCCAACCGGGAGG
CCCTAGGCCTAGAATCGCTGCCAGTACCAAGTTGCCAGTCATCACTCCGGTGCCTCC
ValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAlaAla
5101 ATGTCGTGTGCTGCTCAATGTCTACTCTTGGACAGGCGCACTCGTCACCCGTGCCG
TACAGCACACGACGAGTTACAGAATGAGAACCTGTCCCGCTGAGCAGTGGGCACGCGGC
GluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsnLeu
5161 CGGAAGAACAGAAACTGCCCATCAATGCACTAAGCAACTCGTTGCTACGTCAACCACAATT
GCCTTGTCTTGACGGGTAGTTACGTGATTGAGCAACGATGCAGTGGTGTAA
ValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAspArg
5221 TGGTGTATTCCACCACCTCACGCACTGCTTGCAAAGGCAGAACGAAAGTCACATTGACA
ACCACATAAGGTGGAGTGCACGACGGTTCCGTCTTCAGTGTAAACTGT
LeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAlaSer
5281 GACTGCAAGTTCTGGACAGCCATTACCAAGGACGTACTCAAGGAGGTAAAGCAGCGCGT
CTGACGTTCAAGACCTGTCGTAATGGCCTGCATGAGTCCCTCCAATTGTCGCCGCA
LysValLysAlaAsnLeu
5341 CAAAAGTGAAGGCTAACTTG
GTTTCACTTCCGATTGAAC

FIG. 26-6

FIG. 27 Translation of DNA 12f

IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn
 1 CCATATTTAAATCAGGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAGCTGCCTGCA
 GGTATAAATTAGTCCTACATGCACCCCTCCCCAGCTTGTCCGACCTCGACGGACGT
 TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu
 61 ACTGGACGCCGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCGT
 TGACCTGCGCCCCGCTTGCAACGCTAGACCTCTGTCCCTGTCCAGGCTCGAGTCGGCA
 LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu
 121 TACTGCTGACCACTACACAGTGGCAGGTCTCCGTTCCTCACAAACCCTACCAGCCT
 ATGACGACTGGTATGTCACCGTCCAGGAGGGACAAGGAAGTGTGGATGGTCGGA
 SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal
 181 TGTCACCGGCCCTCATCCACCTCCACCAGAACATTGTGGACGTGCACTTGTACGGGG
 ACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAAACACCTGCACGTCAACATGCC
 GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeu
 241 TGGGGTCAAGCATCGCGCCTGGGCCATTAAGTGGGAGTACGTGTTCTCTGTTCC
 ACCCCAGTCGTAGCGCAGGACCCGGTAATTCACCCATGCAGCAAGAGGACAAGGAAG
 LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu
 301 TGCTTGAGACGCGCGCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAGCG
 ACGAACGTCTGCGCGCGCAGACGAGGACAAACCTACTACGATGAGTATAGGGTTCGCC
 -----Overlap with 14i-----
 AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu
 361 AGGCAGGCTTGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCCGGACGCACGGTC
 TCCGCCGAAACCTTGGAGCATTATGAATTACGTCGTAGGGACCAGGCCCTGCGTGCCAG

 Val
 421 TTGTATC
 AACATAG

FIG. 28 Translation of DNA 35f

-----Overlap with 39c-----

LeuLysGluValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGlu
 1 TGCTCAAGGAGGTTAAAGCAGCGCGTCAAAGTGAAGGCTAACCTGCTATCCGTAGAGG
 ACGAGTCCCTCCAATTCTCGTCGCCGCAGTTCACTTCGATTGAACGATAGGCATCTCC

AlaCysSerLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAsp
 61 AAGCTTGCGAGCCTGACGCCACACTCAGCAAATCCAAGTTGGTTATGGGGCAAAAG
 TTCGAACGTCGGACTGCGGGGTGTGAGTCGGTTAGGTTCAAACCAATACCCGTTTC

ValArgCysHisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAspLeuLeu
 121 ACGTCCGTTGCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTC
 TGCAGGCAACGGTACGGTCTTCCGGCATTGGGTGTAGTTGAGGCACACCTTCTGGAAG

GluAspAsnValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCysVal
 181 TGGAAAGACAATGTAACACCAATAGACACTACCACATGGCTAAAGAACGAGGTTTCTGCG
 ACCTTCTGTTACATTGTGGTTATCTGTGATGGTAGTACCGATTCTGCTCCAAAAGACGC

GlnProGluLysGlyArgLysProAlaArgLeuIleValPheProAspLeuGlyVal
 241 TTCAGCCTGAGAAGGGGGTCGTAAGCCAGCTCGTCTCATCGTGTCCCAGATCTGGCG
 AAGTCGGACTCTCCCCCAGATTGGTCGAGCAGAGTAGCACAAAGGGCTAGACCCGC

ArgValCysGluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAlaValMet
 301 TGCAGCTGTGCGAAAAGATGGCTTGTACGACGTGGTTACAAAGCTCCCTGGCCGTGA
 ACAGCAGCACACGCTTTCTACCGAAACATGCTGCACCAATGTTCGAGGGGAACCGGGCACT

GlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValGlnAla
 361 TGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCTCGTGCAAG
 ACCCTTCGAGGATGCCATAAGGTTATGAGTGGCTGTGCCAACTTAAGGAGCACGTTTC

TrpLysSerLysLysThrProMetGlyPheSerTyrAspThrArgCysPheAspSerThr
 421 CGTGGAAAGTCCAAGAAAACCCAAATGGGGTTCTCGTATGATAACCGCTGCTTGACTCCA
 GCACCTTCAGGTTCTTGAGGTTACCCAAGAGCATACTATGGCGACGAAACTGAGGT

ValThrGluSerAspIleArgThrGluGluAla
 481 CAGTCACTGAGAGCGACATCCGTACGGAGGAGGCA
 GTCAGTGAECTCGCTGTAGGCATGCCTCCTCCGT

FIG. 29 Translation of DNA 19g

1 GluPheLeuValGlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAspThr
 GAATTCCCTCGTGCAAGCGTGGAAAGTCCAAGAAAACCCCAATGGGGTTCTGTATGATACC
 CTTAAGGAGCACGTTCGCACCTCAGGTCTTTGGGGTTACCCCAAGAGCATACTATGG

-----Overlap with 35f-----

61 ArgCysPheAspSerThrValThrGluSerAspIleArgThrGluGluAlaIleTyrGln
 CGCTGCTTGACTCACAGTCAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTACCAA
 GCGACGAAACTGAGGTGTAGTGACTCTCGCTGTAGGCATGCCCTCCGTTAGATGGTT

121 CysCysAspLeuAspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeuTyr
 TGTTGTGACCTCGACCCCCAAGCCCCGCTGGCCATCAAGTCCCTCACCGAGAGGCTTTAT
 ACAACACTGGAGCTGGGGGTTGGGCGCACCGTAGTTCAAGGGAGTGGCTCTCCGAAATA

181 ValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArgAla
 GTTGGGGGCCCTCTACCAATTCAAGGGGGGAGAACACTGCGGCTATCGCAGGTGCCGCGCG
 CAACCCCCGGGAGAATGGTTAAGTCCCCCTCTTGACGCCATAGCGTCCACGGCGCGC

241 SerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArgAla
 AGCGGCGTACTGACAACIAGCTGTGGTAACACCCTACTGCTACATCAAGGCCCGGGCA
 TCGCCGCATGACTGTTGATCGACACCATTGTGGGAGTGAACGATGTAGTTCCGGGCCGT

301 AlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeuVal
 GCCTGTCGAGCCGCAGGGCTCCAGGACTGCAACCATGCTGTTGTCAGGACTTAGTC
 CGGACAGCTCGCGTCCGAGGTCCCTGACGTGGTACGAGCACACACCGCTGCTGAATCAG

361 ValIleCysGluSerAlaGlyValGlnGluAspAlaAla
 GTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGCGAG
 CAATAGACACTTCGCGCCCCCAGGTCCCTGCGCGCGCTC

FIG. 30 Translation of DNA 26g

1 GlyGlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCys
 GGGGGGGAGAACTGCGGCTATCGCAGGTGCCGCAGCAGCGTACTGACAACTAGCTGT
 CCCCCCTCTTGACGCCGATAGCGTCCACGGCGCGTCGCCGCATGACTGTTGATCGACA

61 GlyAsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGln
 GTTAACACCCTCACCTGTTACATCAAGGCCGAGCAGCCTGTCGAGCCGAGGGCTCCAG
 CCATTGAGTGAACATGTAGTTCCGGGCTCGTCGGACAGCTCGCGTCCCAGGTC

-----Overlap with 19g-----

121 AspCysThrMetLeuValCysGlyAspAspLeuValValIleCysGluSerAlaGlyVal
 GACTGCACCATGCTCGTGTGGCGACGACTAGTCGTTATCTGTGAAAGCGCGGGGGTC
 CTGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCAG

181 GlnGluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaPro
 CAGGAGGACGCCGAGCCTGAGAGCCTTCACGGAGGGCTATGACCAGGTACTCCGCC
 GTCCCTCGCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGTCCATGAGGCGGGGG

241 ProGlyAspProProGlnProGluTyrAspLeuGluIleThrSerCysSerSerAsn
 CCTGGGGACCCCCCACAACCAGAATACGACTTGGAGCTCATACATCATGCTCCTCC
 GGACCCCTGGGGGTGTTGGTCTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTG

301 ValSerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThr
 GTGTCACTGCCCCACGACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCC
 ACAGTCAGCGGGTGTGCCCGACCTTCTCCAGATGATGGAGTGGACTGGGACTGGGATGT

361 ThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeu
 ACCCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGAACACTCCAGTCATTCTGGCTA
 TGGGGGGAGCGCTCTCGACGCCACCCCTCTGCGTTCTGTGAGGTCAAGGACCGAT

421 GlyAsnIleIleMetPheAlaProThrLeuTrpAla
 GCGAACATAATCATGTTGCCACACTGTGGCG
 CCGTTGTTAGTACAAACGGGGTGTGACACCCGC

FIG. 31 Translation of DNA 15e

1 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla
 CGCGCTGGAAAGAGGGTCTACTACCTCACCGTGACCCCTACAACCCCCCTCGCGAGAGC
 GCCGCACCTTCCTCCAGATGATGGAGTGGACTGGGATGTTGGGGAGCGCTCTCG

-----Overlap with 26g-----

61 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe
 TGCCTGGAGACAGCAAGAACACTCCAGTCATTCTGGCTAGGCAACATAATCATGTT
 ACGCACCCCTCTGCGTTCTGTGAGGTCAAGGACCGATCCGTTGAGTACAA

121 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla
 TGCCCCCACACTGTGGCGAGGATGATACTGATGACCCATTCTTAGCGTCCTTATAGC
 ACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGTAAAGAAATCGCAGGAATATCG

181 ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu
 CAGGGACCAGCTTGAACAGGCCCTGATGCGAGATCTACGGGGCCIGCTACTCCATAGA
 GTCCCTGGTCGAACCTGTCCGGAGCTAACGCTCTAGATGCCCGGACGATGAGGTATCT

241 ProLeuAspLeuProProIleIleGlnArgLeu
 ACCACTTGATCTACCTCCAATCATTCAAAGACTC
 TGGTGAAGTGGAGGTTAGTAAGTTCTGAG

FIG. 32-1 COMBINED ORF OF DNAs 12f through 15e

IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn
 1 CCATATTTAAATCAGGATGTACTGGGAGGGTCGAACACAGGCAGCTGGAAAGCTGCCCTGCA
 GGTATAAATTAGTCCTACATGCACCCCTCCCCAGCTTGTGTCGACCTTCGACGGACGT
 TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu
 61 ACTGGACGCCGGCGAACGTTGCCATCTGGAAAGACAGGGACAGGTCAGCTCAGCCCCGT
 TGACCTGCCCTCGAACGCTAGACCTCTGTCCCTGTCCAGGCTCGAGTCGGGCA
 LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu
 121 TACTGCTGACCACTACACAGTCAGGTCCTCCGTGTCCTCACAAACCCTACAGCCT
 ATGACGACTGGTGATGTCACCCTCAGGAGGGACAAGGAAGTGTGGATGGTCGGA
 SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal
 181 TGTCCACCGGCCTCATCCACCTCCACCAGAACATTGTCAGTGCAGTACTTGTACGGGG
 ACAGGGTGGCCGGAGTAGGTGGAGGTGGTCTGTAACACCTGCACGTCATGAACATGCC
 GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValLeuLeuPheLeuLeu
 241 TGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGAGTACGTCGTTCTGTCCTTC
 ACCCCAGTCGTAGCGCAGGACCCGGTAATTCAACCTCATGCAGCAAGAGGACAAGGAAG
 LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu
 301 TGCTTGCAGACGCGCGCTGCTCTGCTTGATGCTACTCATATCCAAGCG
 ACGAACGCTGCGCGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCC
 AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu
 361 AGGCGGCTTGGAGAACCTCGTAATACTTAATGCAAGCATCCCTGCCGGACGCACGGTC
 TCCGCCAACCTCTGGAGCATTAATGCACTGTCGTAGGGACCAGGCCCTGCGTGCCAG
 ValSerPheLeuValPhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGly
 421 TTGTATCCTCTCGTGTCTCTGCTTGCATGGTATTGAAGGGTAAGTGGGTGCC
 AACATAGGAAGGAGCACAAGAACGAAACGTACCAACTTCCCATTCAACCCACGGGC
 AlaValTyrThrPheTyrGlyMetTrpProLeuLeuLeuLeuLeuAlaLeuProGln
 481 GAGCGGTCTACACCTCTACGGATGTGGCTCTCCTCTGCTCTGTTGGCGTTGCC
 CTCGCCAGATGTGGAAAGATGCCCTACACGGAGAGGAGGACGAGGACAACCGCAACGGGG
 ArgAlaTyrAlaLeuAspThrGluValAlaAlaSerCysGlyGlyValValLeuValGly
 541 AGCGGGCGTACGCGCTGGACACGGAGGTGGCGCGTGTGTCGTTGGCGTTCTCGTC
 TCGCCCGCATGCGCACCTGTGCCCTCACCGCGCAGCACACCGCCACAACAAGAGCAGC
 LeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrp
 601 GGTTGATGGCGTGTACTCTGACCATATAACAAGCGCTATATCAGCTGGTGCTGTGGT
 CCAACTACCGCGACTGAGACAGTGGTATAATGTCGCGATATAGTCGACCACGAACACCA
 LeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsn
 661 GGCTTCAGTATTTCTGACCAAGTGGAAAGCGCAACTGCACGTGTGGATTCCCCCTCA
 CGAAGTCATAAAAGACTGGTCTCACCTCGCGTGCAGTGCACACGACATGTGGCTGAGACC
 ValArgGlyGlyArgAspAlaValIleLeuLeuMetCysAlaValHisProThrLeuVal
 721 ACGTCCGAGGGGGCGCGACGCCGTACTTACTCATGTGTGCTGTACACCGACTCTGG
 TGCAGGCTCCCCCGCGCTGCAGTAGAATGAGTACACGACATGTGGCTGAGGACC
 PheAspIleThrLysLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAlaSer
 781 TATTGACATACCAAATTGCTGCTGGCGTCTCGAACCCCTTGGATTCTCAAGCCA
 ATAAACTGTAGTGGTTAACGACGACCGGAGAACGCTGGGAAACCTAAGAAGTCGGT
 LeuLeuLysValProTyrPheValArgValGlnGlyLeuLeuArgPheCysAlaLeuAla
 841 GTTGCTAAAGTACCCCTACTTGTGGCGTCCAAGGCCTCTCGGTTCTGCGCGTTAG
 CAAACGAATTTCATGGGATGAAACACGCGCAGGTTCCGGAAAGAGGCCAGACGCGCAATC
 ArgLysMetIleGlyGlyHisTyrValGlnMetValIleIleLysLeuGlyAlaLeuThr
 901 CGCGGAAGATGATCGGAGGCCATTACGTGCAAATGGTCATCTTAAGTTAGGGCGCTTA
 GCGCCCTCTACTAGCCTCCGGTAATGCACGTTACCAAGTAGTAATTCAATCCCCGCGAAT

GlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArg
 961 CTGGCACCTATGTTATAACCATCTCACTCCTCTCGGGACTGGCGCACAAAGGCTG
 GACCGTGGATACAATATTGGTAGACTGAGGAGAACCCCTGACCCGCGTGTGCCGAACG

 AspLeuAlaValAlaValGluProValValPheSerGlnMetGluThrLysLeuIleThr
 1021 GAGATCTGGCGTGGCTGTAGAGCCAGTCGCTTCCTCCAAATGGAGACCAAGCTCATCA
 CTCTAGACCGGCACCGACATCTGGTCAGCAGAAGAGGGTTACCTCTGGTCAGTAGT

 TrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArg
 1081 CGTGGGGGGCAGATACCGCCGCGTGCAGTGCATCAACGCTTGCGCTGTTCCGCC
 GCACCCCCCGTCTATGGCGCGCACGCCACTGTAGTAGTTGCCAACGGACAAAGGCCGG

 ArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeu
 1141 GCAGGGGCCGGAGATACTGCTCGGCCAGCCGATGGAATGGCTCCAAGGGGTGGAGGT
 CGTCCCCGGCCCTCTATGACGAGCCCCTGGCTACCTTACCAAGAGGTTCCCCACCTCCA

 LeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr
 1201 TGCTGGCGCCATCACGGCGTACGCCAGCAGACAAGGGGCCCTAGGGTGCATAATCA
 ACGACCGCGGGTAGTGCCTGCATGCCGCTGTCTCCCCGGAGGATCCCACGTATTAGT

 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla
 1261 CCAGCCTAACTGGCCGGACAAAAACCAAGTGGAGGGTAGGTCCAGATTGTGTCAACTG
 GGTCGGATTGACCGGCCCTGTTGGTCACCTCCACTCCAGGTCTAACACAGTTGAC

 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla
 1321 CTGCCAACCTTCCTGGCAACGTGCATCAATGGGGTGTGGACTGTCACCACGGGG
 GACGGGTTGGAAAGGACCGTGCACGTAGTTACCCACACGACCTGACAGATGGTCCCC

 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp
 1381 CCGGAACGAGGACCATCGCTCACCAAGGGCTGTCACTCCAGATGTATACCAATGTAG
 GCCCTGCTCCTGGTAGCGCAGTGGTCCCAGGACAGTAGGTCTACATATGGTACATC

 GlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys
 1441 ACCAACGACCTTGTGGCTGGCCGCTCCGCAAGGTAGCCGCTATIGACACCCCTGCAC
 TGGTTCTGAAACACCCGACCGGGCGAGCGTCCATCGCGAGTAACTGTGGGACGTGAA

 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg
 1501 GCGGCCTCTCGGACCTTACCTGGTCACGAGGCACGCCATGTCATTCCGTGCGCCCG
 CGCCGAGGAGCCTGAAATGGACCACTGCTCCGTGCGGCTACAGTAAGGCACGCCCG

 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer
 1561 GGGGTGATAGCAGGGGGCAGCCTGCTGCCCCGGCCATTCCCTACTTGAAAGGCTCT
 CCCCCACTATCGTCCCCGTGGACAGCAGCGGGCCGGTAAAGGATGAACCTTCCGAGGA

 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys
 1621 CGGGGGTCCGCTGTTGTGCCCGCGGGCACGCCGTGGCATATTAGGGCCGCGGTGT
 GCCCCCAGGCACACACGGGGCGCCCGTGGCGACCCGTATAAATCCGGCGCCACA

 ThrArgGlyValAlaLysAlaValAspPheIleProValGluAsnLeuGluThrThrMet
 1681 GCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCA
 CGTGGGCACCTCACCGATTCCGCCACCTGAAATAGGGACACCTCTTGGATCTGTGGT

 ArgSerProValPheThrAspAsnSerSerProProValValProGlnSerPheGlnVal
 1741 TGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCACCACTAGTAGTGCCTCAGAGCTCCAG
 ACTCCAGGGGCCACAAGTGCCTATTGAGGAGAGGTGGTCATCACGGGTCTCGAAGGTCC

 AlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAla
 1801 TGGCTCACCTCATGCTCCACAGGCAGCGGCAAAGCACCAAGGTCCCGCTGCATATG
 ACCGAGTGGAGGTACGAGGGTGTCCGTGCCGTTTGTGGTTCCAGGGCGACGTATAC

 AlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGly
 1861 CAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCCTGTGCTGCAACACTGGCTTGT
 GTCGAGTCCCGATATTCCACGATCATGAGTTGGGAGACAACGACGTTGTGACCCGAAAC

AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle

1921 GTGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACCAGGGGTGAGAACAA
 CACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGAGTCCTGGCCCCACTCTGTT
 ThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
 1981 TTACCACTGGCAGCCCCATCACGTACTCCACCTACGGCAAGTTCTGCCGACGGCGGGT
 AATGGTGACCGTCGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAACGGCTGCCGCCA
 SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspAlaThrSer
 2041 GCTCGGGGGCGCTTATGACATAATAATTGTGACGAGTGCCACTCACGGATGCCACAT
 CGAGCCCCCGCGAATACTGTATTAAACACTGCTCACGGTAGGTGCCACGGTGT
 IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
 2101 CCATCTGGGCATCGGCACTGCTTGACCAAGCAGAGACTGCGGGGGCGAGACTGGTTG
 GGTAGAACCCGTAGCCGTACAGGAACTGGTGTCTGACGCCCGCTTGACCAAC
 LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
 2161 TGCTCGCCACCGCCACCCCTCCGGCTCCGTACTGTGCCCATCCAACATCGAGGAGG
 ACGAGCGGTGGCGGTGGGGAGGCCGAGGCACTGACACGGGTAGGGTTGAGCTCCTCC
 AlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIle
 2221 TTGCTCTGTCCACCACCGGAGAGATCCCTTTACGGCAAGGCTATCCCCCTCGAAGTAA
 AACGAGACAGGTGGTGGCCTCTAGGGAAAAATGCCGTTCCGATAGGGGGAGCTTCATT
 LysGlyGlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAla
 2281 TCAAGGGGGGAGACATCTCATCTCTGTCAATTCAAAGAAGTGCGACGAACTCGCCG
 AGTCCCCCCCCTCTGTAGAGTACAAGACAGTAAGTTCTCACGCTGCTTGAGCGGC
 LysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
 2341 CAAAGCTGGTCGCATTGGCATCAATGCCGTGGCTACTACCGCGGTCTGACGTGTCG
 GTTCGACCAAGCGTAACCCGTAGTTACGGCACCGGATGATGGGCCAGAACATGCACAGGC
 IleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGlyTyrThr
 2401 TCATCCCGACCAGCGCGATGTTGTCGTGGCAACCGATGCCCTCATGACCGGCTATA
 AGTAGGGCTGGTCGCCGCTACAACAGCAGCACCGTTGGCTACGGAGTACTGGCCGATAT
 GlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrValAspPheSer
 2461 CCGCGACTTCGACTCGGTGATAGACTGCAATACGTGTCACCCAGACAGTCGATTCA
 GGCGCTGAAGCTGACCACTATCTGACGTTATGCACACAGTGGGCTGTCACTAAAGT
 LeuAspProThrPheThrIleGluThrIleLeuProGlnAspAlaValSerArgThr
 2521 GCCTTGACCCCTACCTTCACCATTGAGACAATCACGCTCCCCCAGGATGCTGCTCCGCA
 CGGAACCTGGGATGGAAGTGGTAACCTGTTAGTGCAGGGGTCTACGACAGAGGGCGT
 GlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGly
 2581 CTCAACGTCGGGCAAGGACTGGCAGGGGAGGCCAGGCATCTACAGATTGTGGCACCGG
 GAGTTGCAGCCCCGTCCTGACCGTCCCCCTCGTCCGTAGATGCTAAACACCGTGGCC
 GluArgProSerGlyMetPheAspSerValLeuCysGluCysTyrAspAlaGlyCys
 2641 GGGAGCCTCCCGCATGTTGACTCGTCCCTCTGTGAGTGTCTATGACGCAGGGCT
 CCCTCGCGGGAGGCCGTACAAGCTGAGCAGGAGACACTCACGATACTGCGTCCGA
 AlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThr
 2701 GTGCTGGTATGAGCTCACGCCGCCAGACTACAGTTAGGCTACGAGCGTACATGAACA
 CACGAACCATACTCGAGTGCAGGGCGCTGTGATGTCATCCGATGCTCGCAIGTACTIGT
 ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeu
 2761 CCCCCGGGCTCCGTGTGCCAGGACATCTGAATTGGGAGGGCGTCTTACAGGCC
 GGGGCCCCGAAGGGCACACGGTCTGGTAGAACTTAAACCCCTCCCGCAGAAATGTCCGG
 ThrHisIleAspAlaHisPheLeuSerGlnSerGlyGluAsnLeuProTyr
 2821 TCACTCATATAGATGCCCACTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTCCCT
 AGTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTCTCACCCCTCTGGAAAGGAA
 LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProSerTrpAsp
 2881 ACCTGGTAGCGTACCAAGCCACCGTGTGCGTAGGGCTCAAGCCCCCTCCCCATCGTGGG
 TGGACCATCGCATGGTTCGGTGGCACACCGCAGTCGGGAGGGTAGCACCC

FIG. 32-3

2941 GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu
 ACCAGATGTGGAAAGTGTGATTGCCTCAAGCCCACCCCTCATGGCCAACACCCCTGC
 TGGTCTACACCTCACAAACTAACGGAGTTCGGGTGGGAGGTACCCGGTTGTGGGGACG

 3001 TyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle
 TATACAGACTGGCGCTGTTAGAATGAAATCACCTGACGCACCCAGTCACCAAATACA
 ATATGTCAGCCCGACAAAGTCTTACTTAGTGGACTGCGTGGTCAGTGGTTATGT

 3061 MetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly
 TCATGACATGCATGTCGGCCGACCTGGAGGTCGTACAGCACCTGGGTGCTCGTGGCG
 AGTACTGTACGTACAGCCGGTGGACCTCCAGCAGTGCTCGTGGACCCACGAGCAACCGC

 3121 ValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArg
 GCGTCCTGGCTGCTTGGCCCGTATTGCCTGTCACAGGCTGCGTGGTCATAGTGGGCA
 CGCAGGACCGACGAAACCGGCCGATAACGGACAGTTGTCGACGCACCAGTATCACCGT

 3181 ValValLeuSerGlyLysProAlaIleProAspArgGluValLeuTyrArgGluPhe
 GGGTCGTCCTGTCGGGAAGCCGGCAATCATACTGACAGGGAAAGTCCTCTACCGAGAGT
 CCCAGCAGAACAGGCCCTCGGCCGTTAGTATGGACTGTCCTCAGGAGATGGCTCTCA

 3241 AspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAla
 TCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAAGGGATGATGCTCG
 AGCTACTCTACCTCTCACGAGAGTCGTGAATGGCATGTCAGCTCGTCCCTACTACGAGC

 3301 GluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluVal
 CCGAGCAGTTCAAGCAGAAGGCCCTCGGCCCTGCAGACCGCGTCCCGTCAGGCAGAGG
 GGCTCGTCAAGTTCGTCTCGGGAGCCGGAGGACGTCTGGCCGAGGGCAGTCCGTCTCC

 3361 IleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMet
 TTATGCCCTGCTGTCCAGACCAACTGGAAAAACTCGAGACCTCTGGCGAAGCATA
 AATAGCGGGGACGACAGGTCTGGTTGACCGTTTGAGCTCTGGAAAGACCCGCTTCGTAT

 3421 TrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnPro
 TGTGGAACCTTCATCAGTGGATAACAATACTTGGCGGGCTGTCAACGCTGCCCTGGTAACC
 ACACCTGAAGTAGTCACCCATGTTATGAACCGCCGAACAGTTGCGACGGACCATTGG

 3481 AlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGln
 CCGCCATTGCTTCATTGATGGCTTTACAGCTGTCACCAGCCCCTAAACCACTAGCC
 GGCGGTAACGAAGTAACCTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATCGG

 3541 ThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAla
 AAACCCCTCCTCTCAACATATTGGGGGGGTGGGTGGCTGCCAGCTGCCGCCCGGTG
 TTTGGGAGGAGAAGTTGTATAACCCCCCCCACCCACCGACGGGTGAGCGGGGGGCCAC

 3601 AlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGlySerValGlyLeuGly
 CCGCTACTGCCCTGTGGCGCTGGCTAGCTGGCGCCCATCGGAGTGTGGACTGG
 GCGATGACGGAAACACCCCGCACCGAACATCGACCGCGCGGTAGCGTACAACCTGACC

 3661 LysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAla
 GGAAGGTCCTCATAGACATCCTGCGAGGTATGGCGGGCGTGGCGGGAGCTCTGTGG
 CCTTCCAGGAGTACTGTAGGAACGTCCCACCGCGCCCGCACCGCCCTCGAGAACACC

 3721 PheLysIleMetSerGlyGluValProSerThrGluAspLeuValAsnLeuLeuProAla
 CATTCAGATCATGAGCGGTGAGGTCCCTCCACGGAGGACCTGGTCAATCTACTGCCCG
 GTAAAGTTCTAGTACTGCCACTCCAGGGAGGTGCCCTGGACCAAGTTAGATGACGGGC

 3781 IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis
 CCATCCTCTGCCCGAGCCCTCGTAGTCGGCGTGTGCAAGCAATACTGCCCG
 GGTAGGAGAGCGGGCCTCGGAGCATCGCCGACACAGTCGTTATGACGCCCG

 3841 ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg
 ACGTTGGCCCGGGCGAGGGGGCAGTGCAGTGGATGAACCGGCTGATAGCCTCGCCTCCC
 TGCAACCGGGCCCGCTCCCCGTACGTACCTACTTGGCCGACTATCGGAAGCGGAGGG

 GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaArgValThr

FIG. 32-4

3901 GGGGAAACCATGTTCCCCCACCACACTACGTGCCGGAGAGCGATGCAGCTGCCCGCTCA
 CCCCTGGTACAAAGGGGTGCGTGATGCACGGCCTCTCGCTACGTCGACGGGCGCAGT
 AlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSer
 3961 CTGCCATACTCAGCAGCCTCACGTGAACCCAGCTCCTGAGGCGACTGCACCAGTGGATAA
 GACGGTATGAGTCGTCGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTCACCTATT
 SerGluCysThrThrProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCys
 4021 GCTCGGAGTGTACCACTCCATGCTCCGGTCTGGCTAAGGGACATCTGGACTGGATAT
 CGAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGAGACCCCTGACCTATA
 GluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGly
 4081 GCGAGGTGTGAGGCACTTTAACCTGGCTAAAGCTAAGCTCATGCCACAGCTGCCTG
 CGCTCCACAACCTCGCTGAAATTCTGGACCGATTTCGATTGAGTACGGTGTGACGGAC
 IleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMet
 4141 GGATCCCCCTTGTGCTCTGCCAGCGGGTATAAGGGGGTCTGGCAGTGGACGGCATCA
 CCTAGGGAAACACAGGACGGTCGCGCCCATTCCCCCAGACCGCTCACCTGCCGTAGT
 HisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArg
 4201 TGCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAAACGGGACGATGA
 ACGTGTGAGCGACGGTACACCTCGACTCTAGTGACTGTACAGTTTGCCCTGCTACT
 IleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyr
 4261 GGATCGTCGGTCCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTCCCCATTATGCC
 CCTAGCAGCCAGGATCCTGGACGTCCCTGTACACCTCACCCCTGGAAGGGTAATTACGGA
 ThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPheAlaLeuTrpArgVal
 4321 ACACCACGGGCCCCCTGTACCCCCCTTCGCGCGAACTACACGTTCGCGCTATGGAGGG
 TGTGGTGCCTGGGGACATGGGGGAAGGACGCGGCTGATGTGCAAGCGCGATAACCTCCC
 SerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMet
 4381 TGTCTGAGAGGAATATGTGGAGATAAGGCAGGTGGGGACTTCACTACGTGACGGGTA
 ACAGACGTCTCCTATACACCTCTATTCCGTCCACCCCTGAAGGTGATGCACTGCCCAT
 ThrThrAspAsnLeuLysCysProCysGlnValProSerProGluPhePheThrGluLeu
 4441 TGACTACTGACAATCTCAAATGCCGTGCCAGGTCCCATGCCGAATTTCACAGAAT
 ACTGATGACTGTTAGAGTTACGGGCACGGTCCAGGGTAGCGGGCTAAAAAGTGTCTTA
 AspGlyValArgLeuHisArgPheAlaProProCysLysProLeuLeuArgGluGluVal
 4501 TGGACGGGTGCGCCTACATAGGTTGCCAGGTCCCATGCCGAATTGCTGCGGGAGGAGG
 ACCTGCCAACGCCATGTATCCAAACGCCGGGGACGTTGGAAACGACGCCCTCC
 SerPheArgValGlyLeuHisGluTyrProValGlySerGlnLeuProCysGluProGlu
 4561 TATCATCAGAGTAGGACTCCACGAATAACCGGTAGGGTCGCAATTACCTTGCGAGCCCG
 ATAGTAAGTCTCATCCTGAGGTGCTTATGGGCATCCCAGCGTTAATGGAACGCTCGGGC
 ProAspValAlaValLeuThrSerMetLeuThrAspProSerHisIleThrAlaGluAla
 4621 AACCGGACGTGGCGTGTGACGTCCATGCTACTGATCCCTCCATATAACAGCAGAGG
 TTGGCCTGCACCGGCACAACTGCAGGTACGAGTGACTAGGGAGGGTATATGTCGTC
 AlaGlyArgArgLeuAlaArgGlySerProProSerValAlaSerSerAlaSerGln
 4681 CGGGCGGGCGAAGGTGGCGAGGGATACCCCCCTCTGTGGCCAGCTCTCGGCTAGCC
 GCCGGCCCGCTTCAACCGCTCCCTAGTGGGGGAGACACCGGTCGAGGAGCCGATCGG
 LeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHisAspSerProAspAlaGlu
 4741 AGCTATCCGCTCCATCTCTCAAGGCAACTTGACCGCTAACCATGACTCCCTGATGCTG
 TCGATAGGCGAGGTAGAGAGTTCCGTTGAACGTGGCGATTGGTACTGAGGGAGGGACTACGAC
 LeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGlu
 4801 AGCTCATAGAGGCCAACCTCTATGGAGGCAGGAGATGGCGCAACATCACCAGGGTTG
 TCGAGTATCTCCGGTTGGAGGATACCTCCGCTTACCCGCCGTTGAGTGGTCCCAAC
 SerGluAsnLysValValIleLeuAspSerPheAspProLeuValAlaGluGluAspGlu
 4861 AGTCAGAAAACAAAGTGGTATTCTGGACTCCTCGATCCGCTTGTGGCGGAGGAGGACG
 TCAGTCTTGTTCACCACTAACGACCTGAGGAAGCTAGGCGAACACCGCCTCCCTG

FIG. 32-5

2212511

4921 ArgGluIleSerValProAlaGluIleLeuArgLysSerArgArgPheAlaGlnAlaLeu
 AGCGGGAGATCTCGTACCCGCAGAAATCCTGCGGAAGTCTCGGAGATTGCCAGGCC
 TCGCCCTCTAGAGGCATGGCGTCTTAGGACGCCCTCAGAGCCTCTAACGGGTCCGGG

 4981 ProValTrpAlaArgProAspTyrAsnProProLeuValGluThrTrpLysLysProAsp
 TGCCCCGTTGGCGCGGCCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAAGCCG
 ACGGGCAAACCCGCGCCGCCCTGATATTGGGGCGATCACCTCTGCACCTTTTCGGG

 5041 TyrGluProProValValHisGlyCysProLeuProProLysSerProProValPro
 ACTACGAACCACCTGTGGTCCATGGCTGTCCACCTCAAAGTCCCCTCTGTGC
 TGATGCTTGGTGACACCAGGTACCGACAGGCGAAGGTGGAGGTTCAAGGGAGGACACG

 5101 ProProArgLysLysArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAla
 CTCCGCCTCGGAAGAAGCGGACGGTGGCCTCACTGAATCAACCTATCTACTGCCTTGG
 GAGGGGAGCCTCTCGCCTGCCACCAGGAGTGAACTTAGTGGATAGATGACGGAACC

 5161 GluLeuAlaThrArgSerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThr
 CCGAGCTGCCACCAGAACGCTTGGCAGCTCCTCAACTCCGGCATTACGGGCGACAATA
 GGCTCGAGCGGTGGTCTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCGCTGTTAT

 5221 ThrThrSerSerGluProAlaProSerGlyCysProProAspSerAspAlaGluSerTyr
 CGACAACATCCTCTGAGCCGCCCTCTGGCTGCCCGGACTCCGACGCTGAGTCCT
 GCTGTTGAGGAGACTCGGGCGGGGAGACCGACGGGGGGCTGAGGCTGCGACTCAGGA

 5281 SerSerMetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrp
 ATTCCCTCATGCCCTGGAGGGGAGCCTGGGATCCGGATCTIAGCGACGGTCAT
 TAAGGAGGTACGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAATCGCTGCCAGTA

 5341 SerThrValSerSerGluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSer
 GGTCAACGGTCAGTAGTGAGGCCAACCGGGAGGATGTCGTGCTCAATGTCTACT
 CCAGTTGCCAGTCATCACTCCGGTTGCCCTCACAGCACACGACGAGTTACAGAATGA

 5401 TrpThrGlyAlaLeuValThrProCysAlaAlaGluGlnLysLeuProIleAsnAla
 CTGGACAGGCCACTCGTACCCGCGCCGGAAGAACAGAAACTGCCATCAATG
 GAACCTGTCGCGTGAGCAGTGGGACCGCCTTGTGACGGGTAGTTAC

 5461 LeuSerAsnSerLeuLeuArgHisHisAsnLeuValTyrSerThrSerArgSerAla
 CACTAAGCAACTCGTTGCTACGTCACCACAATTGGGTGATTCCACCTCACGCACTG
 GTGATTGAGCAACGATGCAAGTGGTAAACACATAAGGTGGAGTGCAC

 5521 CysGlnArgGlnLysLysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGln
 CTTGCCAAAGGCCAGAACGAAAGTCACATTGACAGACTGCAAGTTCTGGACAGCCATTAC
 GAACGGTTCCGCTTCTCAGTGTAAACTGTCAGCTCAAGACCTGCGTAATGG

 5581 AspValLeuLysGluValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerVal
 AGGACGTACTCAAGGAGGTAAAGCAGCGCGTCAAAAGTGAAGGCTAATTGCTATCCG
 TCCTGCATGAGTCCCTCCAATTGTCGCGCGAGTTTCACTCCGATTGAACGATAGGC

 5641 GluGluAlaCysSerLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAla
 TAGAGGAAGCTTGAGCCCTGACGCCACACTCAGCAAATCCAAGTTGGTTATGGGG
 ATCTCCTCGAACGTCGGACTGCGGGGTGTGAGTCGGTTAGGTTCAAACCAATACCC

 5701 LysAspValArgCysHisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAsp
 CAAAAGACGTCGTTGCCATGCCAGAACGCCGTAACCCACATCAACTCCGTGTGGAAAG
 GTTTCTGCAGGCAACGGTACGGCTTCCGGCATGGGTGTAGTTGAGGCACACCTTC

 5761 LeuLeuGluAspAsnValThrProIleAspThrThrIleMetAlaLysAsnGluValPhe
 ACCTTCTGGAAGACAATGTAACACCAATAGACACTACCATCATGGCTAAGAACGAGGTTT
 TGGAGACCTCTGTTACATTGTTGAGGTTATCTGTGATGGTAGTACCGATTGCTCCAAA

 5821 CysValGlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeu
 TCTGCGTTCAAGCCTGAGAACGGGGGTGTAAGCCAGCTCGTCTCATCGTGTCCCCGATC
 AGACGCAAGTCGGACTCTCCCCCAGATTGGTCGAGCAGAGTAGCACAAGGGCTAG

 GlyValArgValCysGluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAla

221'51'

5881 TGGGCGTGCACGCGTGTGCGAAAAGATGGCTTGACGACGTGGTTACAAAGCTCCCCTGG
 ACCCGCACGCGCACACGCTTCTACCGAAACATGCTGACCAATGTTCGAGGGAAACC
 ValMetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuVal
 5941 CCGTGATGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGGGTTGAATTCCCTCG
 GGCACATACCCCTCGAGGATGCCATAAGGTTATGAGTGGCTGTGCCAACCTAAGGAGC
 GlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAspThrArgCysPheAsp
 6001 TGCAAGCGTGGAAAGTCCAAGAAACCCCAATGGGTTCTCGTATGATACCCGCTGCTTTG
 ACGTTCGCACCTTCAGGTTCTTGGGTTACCCCAAGAGCATACTATGGCGACGAAAC
 SerThrValThrGluSerAspIleArgThrGluGluAlaIleTyrGlnCysCysAspLeu
 6061 ACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTACCAATGTTGTGACC
 TGAGGTGTCAGTGACTCTCGTGTAGGCATGCCCTCCGTTAGATGGTTACAACACTGG
 AspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeuTyrValGlyGlyPro
 6121 TCGACCCCCAAGCCCGCGTGGCATCAAGTCCCTCACCGAGAGGCTTATGTTGGGGGCC
 AGCTGGGGGTTGGCGCACCGTAGTTCAAGGAGTGGCTCTCCGAAATAACACCCCCCGG
 LeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeu
 6181 CTCTTACCAATTCAAGGGGGAGAACTGCGGCTATCGCAGGTGCGCGAGCGGGCGTAC
 GAGAATGGTTAAGTCCCCCTTTGACGCCATAGCGTCCACGGCGCTGCCGCATG
 ThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAla
 6241 TGACAACTAGCTGTGGTAACACCCACTTGCTACATCAAGGCCGGCAGCCTGTCGAG
 ACTGTTGATCGACACCATTGTGGAGTAGTGTAGTTCCGGCCGTCGGACAGCTC
 AlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeuValValIleCysGlu
 6301 CCGCAGGGCTCCAGGACTGCACCATGCTCGTGTGGCGACGACTTAGTCGTTATCTGTG
 GGCGTCCCGAGGTCTGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACAC
 SerAlaGlyValGlnGluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArg
 6361 AAAGCGGGGGGTCCAGGAGGACCGCGAGCCTGAGAGCCTCACGGAGGCTATGACCA
 TTTCGCCCCCAGGTCTCGCCGCTCGGACTCTCGGAAGTGCCTCCGATACTGGT
 TyrSerAlaProProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSer
 6421 GGTACTCCGCCCCCCTGGGGACCCCCACAACCAGAAATACGACTTGGAGCTCATAACAT
 CCATGAGGGGGGGGGACCCCTGGGGGTGTTGGTCTTATGCTGAACCTCGAGTATTGTA
 CysSerSerAsnValSerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThr
 6481 CATGCTCTCCAACGTGTCAGTCGCCCCACACGGCGCTGGAAAGAGGGCTACTACCTCA
 GTACGAGGAGGTGACAGTCAGCGGGTGTGCCGCACCTTCTCCAGATGATGGAGT
 ArgAspProThrThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProVal
 6541 CCCGTGACCCCTACAACCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAG
 GGGCACTGGGATGTTGGGGAGCGCTCTGACGCACCCCTGTCGTTCTGTGAGGTC
 AsnSerTrpLeuGlyAsnIleIleMetPheAlaProThrLeuTrpAlaArgMetIleLeu
 6601 TCAATTCTGGCTAGGCAACATAATCATGTTGCCACACTGTGGCGAGGATGATAC
 AGTTAAGGACCGATCCGTTGATTAGTACAAACGGGGTGTGACACCCGCTCTACTATG
 MetThrHisPhePheSerValLeuIleAlaArgAspGlnLeuGluGlnAlaLeuAspCys
 6661 TGATGACCCATTCTTAGCGCTTATAGCCAGGGACAGCTGAAACAGGCCCTCGATT
 ACTACTGGTAAAGAAATCGCAGGAATACGGTCCCTGGTCGAACCTGTGAGGCTAA
 GluIleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuProProIleIleGlnArg
 6721 GCGAGATCTACGGGGCTGCTACTCCATAGAACCACTTGATCTACCTCAATCATTCAA
 CGCTCTAGATGCCCGGACGATGAGGTATCTGGTGAACTAGATGGAGGTTAGTAAGTT
 Leu
 6781 GACTC
 CTGAG

FIG. 32-7

39

2212511

FIG. 33 LEGEND

Lane Number	Chimp Reference Number	Infection Type	Sample date (days) (0=inoculation day)	ALT (alanine) aminotransferase level in sera) mu/ml
1	1	NANB	0	9
2	1	NANB	76	71
3	1	NANB	118	19
4	1	NANB	154	N/A
5	2	NANB	0	5
6	2	NANB	21	52
7	2	NANB	73	13
8	2	NANB	138	N/A
9	3	NANB	0	8
10	3	NANB	43	205
11	3	NANB	53	14
12	3	NANB	159	6
13	4	NANB	-3	11
14	4	NANB	55	132
15	4	NANB	83	N/A
16	4	NANB	140	N/A
17	5	HAV	0	4
18	5	HAV	25	147
19	5	HAV	40	18
20	5	HAV	268	5
21	6	HAV	-8	N/A
22	6	HAV	15	106
23	6	HAV	41	10
24	6	HAV	129	N/A
26	7	HAV	0	7
27	7	IHAV	22	83
28	7	RAV	115	5
29	7	RAV	139	N/A
30	8	HAV	0	15
31	8	HAV	26	130
32	8	HAV	74	8
33	8	RAV	205	5
34	9	HBV	-290	N/A
35	9	HBV	379	9
36	9	HBV	435	6
37	10	HBV	0	8
38	10	HBV	111-118 (pool)	96-156 (pool)
39	10	HBV	205	9
40	10	HBV	240	13
41	11	HBV	0	11
42	11	HBV	28-56 (pool)	8-100 (pool)
43	11	HBV	169	9
44	11	HBV	223	10

2212511

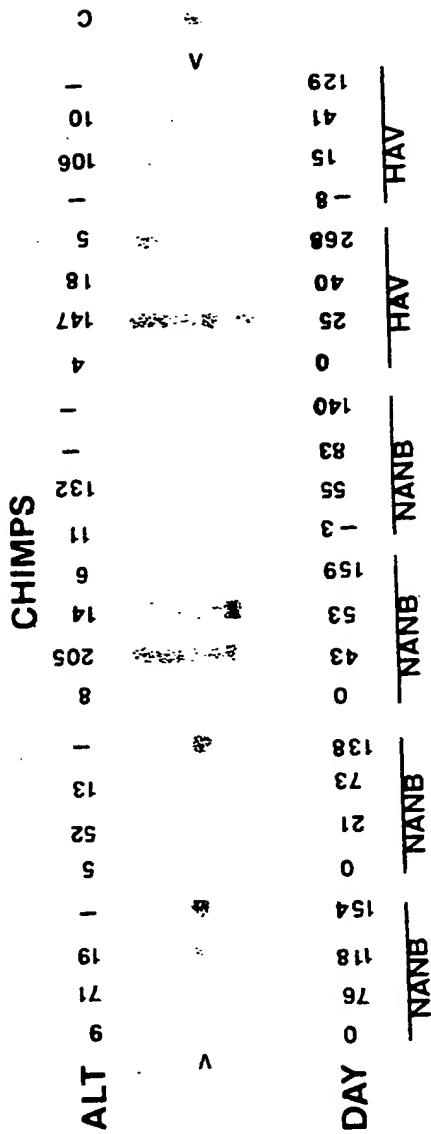


FIG. 33-1

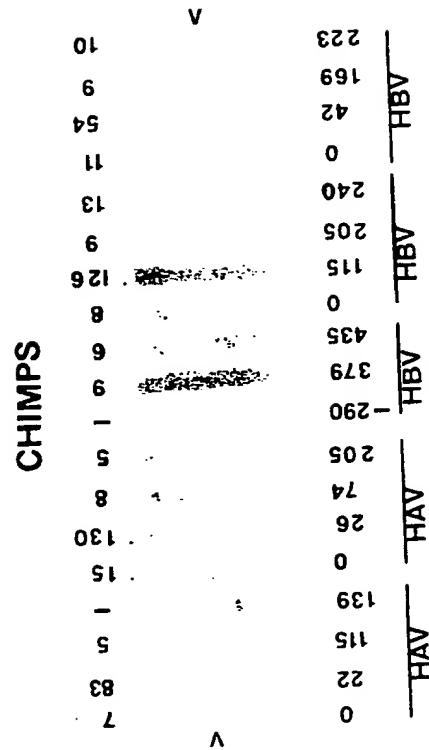


FIG. 33-2

2212511

FIG. 34 LEGEND

Lane Number	Patient Reference Number	Diagnosis	ALT Level (m μ /ml)
1	1 ¹	NANB	1354
2	11	NANB	31
3	21	NANB	14
4	21	NANB	79
5	21	NANB	26
6	31	NANB	78
7	31	NANB	87
8	31	NANB	25
9	41	NANB	60
10	41	NANB	13
11	51	NANB	298
12	51	NANB	101
13	61	NANB	474
14	61	NANB	318
15	71	NANB	20
16	71	NANB	163
17	81	NANB	44
18	81	NANB	50
19	9	NANB	N/A
20	10	NANB	N/A
21	11	NANB	N/A
22	12	Normal	N/A
23	13	Normal	N/A
24	14	Normal	N/A
26	30174	Normal	N/A
27	30105	Normal	N/A
28	30072	Normal	N/A
29	30026	Normal	N/A
30	30146	Normal	N/A
31	30250	Normal	N/A
32	30071	Normal	N/A
33	15	AcuteHAV	N/A
34	16	AcuteHAV	N/A
35	17	AcuteHAV	N/A
36	18	AcuteHAV	N/A
37	48088	AcuteHAV	N/A
38	47288	AcuteHAV	N/A
39	47050	AcuteHAV	N/A
40	46997	AcuteHAV	N/A
41	19	Convalescent HBV	N/A
42	20	(anti-HBSag+ve;	N/A
43	21	anti-HBCag+ve)	N/A
44	22	(anti-HBSag+ve;	N/A
45	23	anti-HBCag+ve)	N/A
46	24	(anti-HBSag+ve;	N/A
47	25	anti-HBCag+ve)	N/A
48	26	(anti-HBSag+ve;	N/A
49	27	anti-HBSag+ve)	N/A

¹ Sequential serum samples were assayed from these patients

42

2212511

FIG. 34-1

FIG. 34-2

20

45

15

40

10

35

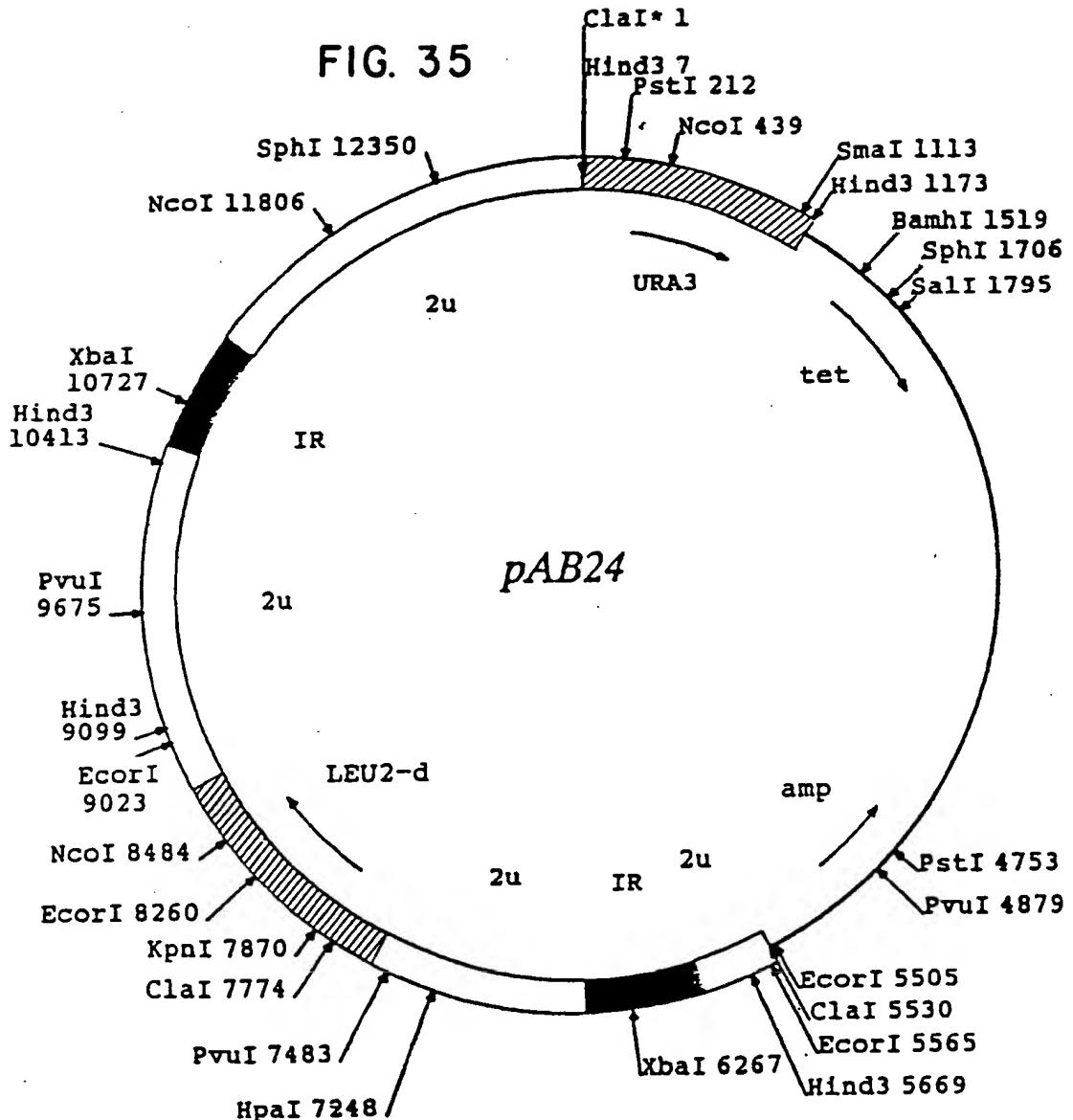
5

30

25

2212511

FIG. 35



2212511

FIG. 36-1 COOH-terminus of SOD-C100 Fusion Polypeptide

221251

CGACCGCGGCGGTAGCCGTACAACCTGACCCCTCCAGGAGTATCTGTAGGAACGTCCC
 TyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValPro
 901 TATGGCGCGGGCGTGGCGGGAGCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCC
 ATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTGCCACTCCAGGGG
 SerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValVal
 961 TCCACGGAGGACCTGGTCAATCTACTGCCGCCATCCTCTGCCCGAGCCCTCGTAGTC
 AGGTGCCTCTGGACCAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCA^G
 GlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGln
 1021 GGC GTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGCGAGGGGGCAGTGCAG
 CCGCACCAAGACACGTCGTTATGACCGGGCGTGCAACCGGGCCGCTCCCCCGTCACGTC
 <<<<<<<<<<< NANBH] [---extra
 TrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProValHisHis
 1081 TGGATGAACCGGCTGATAGCCTCGCCTCCGGGGAAACCATGTTCCCCAGTCCATCAT
 ACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTGGTACAAAGGGTCAGGTAGTA
 -----]
 LysArgOP
 1141 AAGCGTTGACGCTCCCTACGGGTGGACTGTGGAGAGACAGGGCACTGCTAAGGCCAAAT
 TTGCGCAACTGCGAGGGATGCCACCTGACACCTCTGTCCCGTGACGATTCCGGTTA
 1201 CTCAGCCATGCATCGAGGGGTACAATCGTATGCCAACAACTAGCGCGTACGTAAAGTC
 GAGTCGGTACGTAGCTCCCATGTTAGGCATACCGGTTGTGATCGCGATGCAATTTCAG
 1261 TCCTTTCTCGATGGTCCATACCTTAGATGCGTTAGCATTAATCCGAATT
 AGGAAAGAGCTACCAGGTATGGAATCTACGCAATCGTAATTAGGCTTAAG

FIG. 36-2

C

4k

2212511



FIG. 37 a

1 2 3

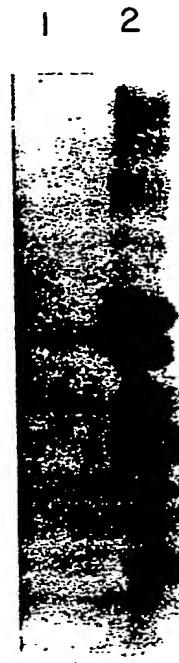


FIG. 37 b

A

FIG. 39

47

2212511

1 2 3 4



FIG. 38

48

2212511

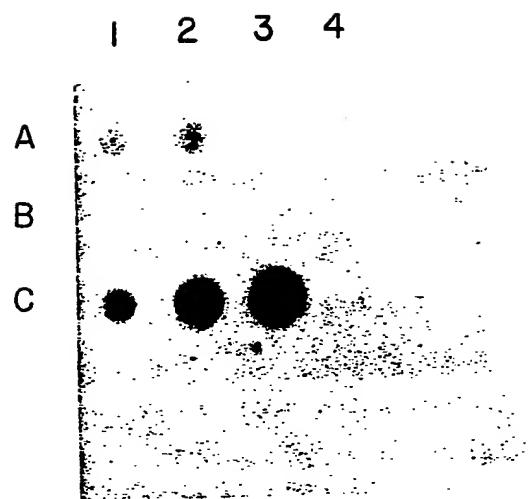


FIG. 40

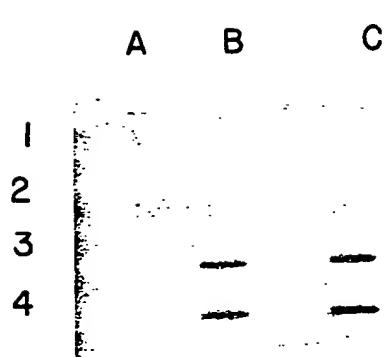


FIG. 41a

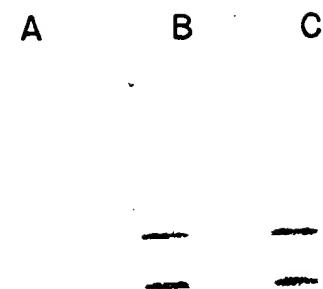


FIG. 41b

2212511

FIG. 4I-I

Homology between the HCV polypeptide encoded by combined ORF of clones 14i through 39c) and the non-structural protein of the Dengue flavivirus(MNWVD1).

	10	20	30	40	50
HCV	EYVVILLFLLLADARVCSCLWMMILLISQAEAALENLVILNAASLAGTHGLVSFLVFFCFA				
MNWVD1	AVSFVTLITGNMSFRDLGRVMVMGATMTDDIGMGVTYLALLAALKVRPTFAAGLLLRLK	130	140	150	160
					170
HCV	60 70 80 90 100 110				
	WYLKGKWPVGAVYTFYGMWP LALPQRAYALDTEVAASC GG VVLVGLMALTLSPYY				
MNWVD1	TSKELMMTTIGIVLLSQSTIPETILELTDALALGMMVLKMVRKMEKYQLAVTMAILCVP	190	200	210	220
					230
HCV	120 130 140 150 160 170				
	KRYISWCWLWQYFLTRVEAQLHVVWIPPLNVRGGRAVILLMCAVHPTLVFDITKLLLAV				
MNWVD1	NAVILQNAWKVSCTILAVVSVSPLFLTSSQQKADWIPLALTIGLNPATAIF-LTTLSRDN	250	260	270	280
					290
HCV	180 190 200 210 220 230				
	FGPLWILQASLLKVPYF-VRVQGLLRF-CALARCKMIGGHYVQMVIIKLGALTGTYVYNHL				
MNWVD1	KKRSWPLNEAIMAVGMVISLASSLLKNDIPMTGPLVAGGLLTVCYV-LTG RSA DLELERA	300	310	320	330
					340
HCV	360 370 380 390 400 410				
	ADV K-WEDQAEISGSSPILSITISE-DGSMSIKNEEEEQTLTILIRTGLLVISG---LFP				
HCV	300 310 320 330 340 350				
	PADGMVKSGWRLLAPITAYAQQT RGLLGCIIITSLTGRDKNQVEGEVQIVSTAATFLATC				
MNWVD1	VSIPI TAAWYLWEVKKQRAGVLWDVPSPPVGKA ELEDGAYRIKQKGILGYSQIGAGVY	420	430	440	450
					460
HCV	360 370 380 390 400 410				
	INGVCWTYHGAGTRTI ASPKG PVIQMYTNVDQDLV---GWPAPQGSRS LTPCTCGSSD				
MNWVD1	KEGTFHTMWHVTRGA VLMHKGKRIEPSWADVKDLVSCGGW KLEG EWKE GEEVQVLALE	480	490	500	510
					520
HCV	420 430 440 450 460 470				
	LYLVTRHADVIPVRRRGDSRG SLLSPRPIS YLKGS SGGPLLCPAGH AVG IFRAAV CTRGV				
MNWVD1	PGKNPRAVQTKPGLF KTN--AGTIGAVSLDFSPGTGS SPIIDKKGKVVG LYNGNVVTRSG	540	550	560	570
					580
HCV	480 490 500 510 520 530				
	AKAVDFIPVENLETTMRSPVFTDNSSPPVVPQSFQVAHLHAPT GSGKS--TKVPAAYAAQ				
MNWVD1	AYVSAIAQTEK--SIEDNPEIE DDIFRK--RKLTIMDLHPGAGKT KRYLPAIVRGAIKR	600	610	620	630
					640
	540 550 560 570 580				

2212511

HCV	GYKVLVLPNS--VAATLGFGAYMSKAHGIDPNIRTGVRTITGGSPITYSTYKFADGGC											
MNWVD1	650	660	670	680	690	700	650	660	670	680	690	700
HCV	590	600	610	620	630	640	590	600	610	620	630	640
MNWVD1	RVPYNLIIMDEAHFTDPASIAARGYISTRVE-MGEAAGIFMTATPPGSRD-PFPQSNAP	710	720	730	740	750	710	720	730	740	750	760
HCV	650	660	670	680	690	700	650	660	670	680	690	700
MNWVD1	ALSTTGEIFPYGKAIPLEVIGGRHLIFCHSKKKCDELAALKVALGINAVAYYRGLDVSV	770	780	790	800	810	770	780	790	800	810	820
HCV	IPTSGDVVVVATDALMTGYTGFDSVIDCNCVTQTVDFSLDPTFTIETITLPQDAVSRT	710	720	730	740	750	760	710	720	730	740	750
MNWVD1	SEYVKTRTNNDWNFVVTTDISEMGANFAERVIDP RRCM KPVILTDGEERVILAGPMPVTH	830	840	850	860	870	880	830	840	850	860	870
HCV	QRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT	770	780	790	800	810	820	770	780	790	800	810
MNWVD1	SS											

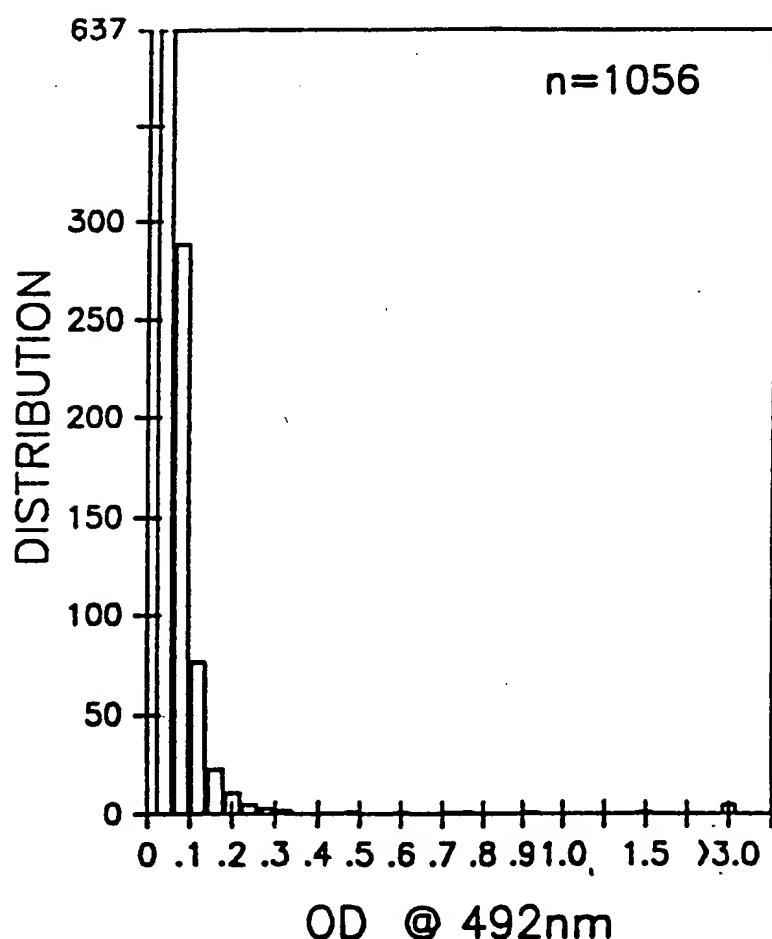
FIG. 41-2

51

2212511

FIG. 43
DISTRIBUTION OF RANDOM SAMPLES

C100-3 Ag ELISA Preclinical Kit
416ng C100/WELL, 2 HRS 37°C, 20ul SAMPLE



C

52

2212511

FIG. 44
Distribution of O.D. Values for
Random Blood Donor Samples Tested with Two ELISA
Configurations
C100-3 Ag ELISA MoAB vs Polyclonal

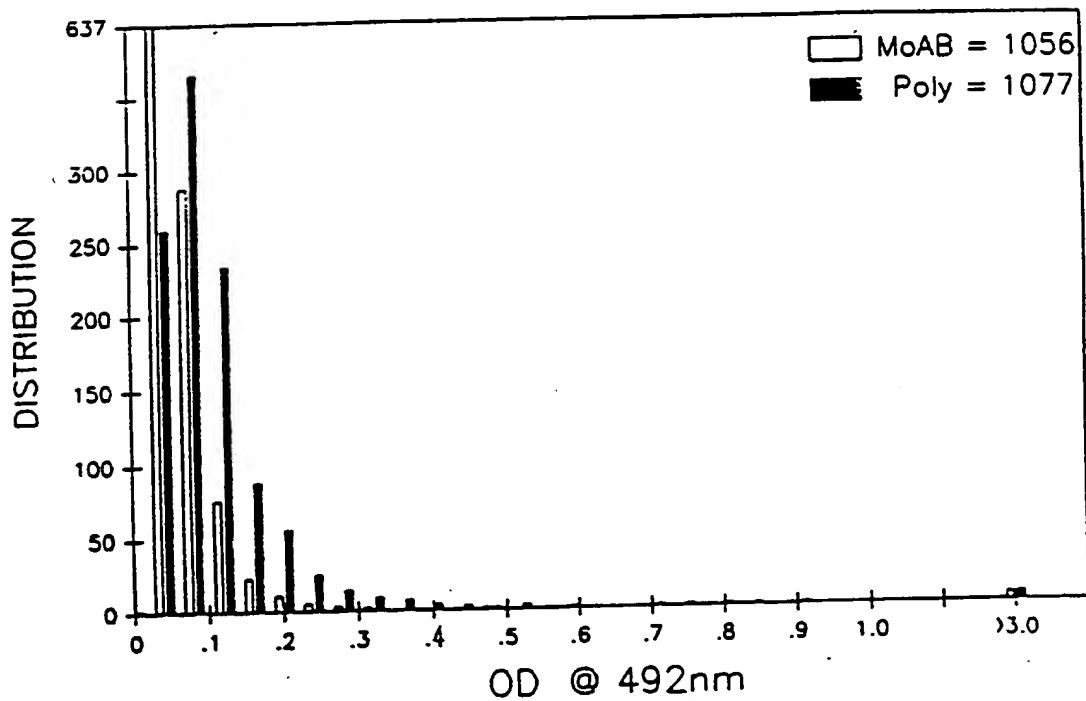


FIG. 45

<u>Name</u>	<u>Common Sequence</u>	<u>Variable Sequence</u>
5'-3-1	AAGCTTGATCGAATTTC	CGATCTTGC
-2		CGATCCTGC
-3		CGATCATGC
-4		CGATCGTGC
-5		CGAAGTTGC
-6		CGAAGCTGC
-7		AGATCTTGC
-8		AGATCCTGC
-9		AGATCATGC
-10		AGATCGTGC
-11		AGAAGTTGC
-12		AGAAGCTGC
-13		CGATCTTGT
-14		CGATCCTGT
-15		CGATCATGT
-16		CGATCGTGT
-17		CGAAGTTGT
-18		CGAAGCTGT
-19		AGATCTTGT
-20		AGATCCTGT
-21		AGATCATGT
-22		AGATCGTGT
-23		AGAAGTTGT
-24		AGAAGCTGT
-25		CGCTCTTGC
-26		CGCTCCTGC
-27		CGCTCATGC
-28		CGCTCGTGC
-29		CGCAGTTGC
-30		CGCAGCTGC
-31		CGCTCTTGT
-32		CGCTCCTGT
-33		CGCTCATGT
-34		CGCTCGTGT
-35		CGCAGTTGT
-36		CGCAGCTGT

2212511

FIG. 46-1 Translation of DNA k9-1

GlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGly
 1 CAGGCTGTCTGAGAGGCTAGCCAGTGCACCCCTTACCGATTGACCAGGGCTGGG
 GTCCGACAGGACTCTCCGATCGTCGACGGCTGGGAATGGCTAAACTGGTCCCACCC

ProIleSerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrPro
 61 GCCCTATCAAGTTATGCCAACGGAGCGGCCCTACTGCTGGCACTACC
 CGGGATAGTCAATACGGTIGCCTCGCCGGGCTGGTCGCGGGGATGACGACCGTGATGG

ProLysProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThr
 121 CCCCAAAACCTTGCCTATTGTGCCCGAGAGTGTGTGGTCCGGTATATTGCTTC
 GGGGTTTGGAACGCCATAACACGGGCGTCTCACACACACCAGGCCATATAACGAAGT

ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
 181 CTCCCAGCCCCGTGGTGGGAACGACCGACAGGTGGCGCGCCCACCTACAGCTGGG
 GAGGGTCGGGCACCACCCCTGCTGGCTGGCAGCCGCGCGGGTGGATGTCGACCC

GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
 241 GTGAAAATGATAACGGACGCTCTGCTTAACAATACCAACTGGCCACCGCTGGGAAATTGGT
 CACTTTACTATGCCTGCAGAACAGGAAATTGTTATGGTCCGGTGGGACCCCTAACCA

GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
 301 TCGGTTGTACCTGGATGAACTCAACTGGATTCAACAAAGTGTGGAGCGCCCTTG
 AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTCACACGCCCTCGCGGAGGAACAC

IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
 361 TCATCGGAGGGGGCGGGCAACAAACACCCCTGCACTGCCCACTGATTGCTCCGCAAGCATC
 AGTAGCCTCCCCGCCCTTGTGGGACGTGACGGGGTGAACAGAAGGCCTCGTAG

AspAlaThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAsp
 421 CGGACGCCACATACTCTGGTGCAGCTCCGGTCCGGTGGATCACACCCAGGTGCCTGGTCG
 GCCTGGGTGTATGAGAGGCCACGCCAGGGACCTAGTGTGGTCCACGGACCAGC

TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
 481 ACTACCCGTATAGGCTTGGCATTATCCTGTACCATCAACTACACTATTTAAATCA
 TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGATAAAATTAGT

MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
 541 GGAIGTACGTGGGAGGGGTCGAGCACAGGTGGAAGCTGCCTGCAACTGGACGCCGGCG
 CCTACATGCACCCCTCCCCAGCTGTCCGACCTCGACGGACGTTGACCTGCCGCCCCCG

ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
 601 AACGTTGCGATCTGGAAAGATAGGGACAGGTCCGAGCTCAGCCGTTACTGCTGACCACTA
 TTGCAACGCTAGACCTTCTATCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGTAT

GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIle
 661 CACAGTGGCAGGTCTCCGTGTCCTCACAAACCTGCCAGCCTGTCCACCGGCCCTCA
 GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTGGGACGGTCGGAACAGGTGGCCGGAGT

Overlap with Combined ORF of DNAs 12f through 15e-----
 HisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAla
 721 TCCACCTCCACCAGAACATTGTGGACGTGCACTGTACGGGGTGGGTCAAGCATCG
 AGGTGGAGGTGGTCTTGTAAACACCTGCACGTCAACATGCCCCACCCAGTTCGTAGC

SerTrpAlaIleLysTrpGluTyrValValLeuLeuLeuLeuLeuAlaAspAlaArg
 781 CGTCCTGGGCCATTAAGTGGGAGTACGTGCTCCTGTTCTGCTGCTGAGACGCC
 GCAGGACCGGTAATTACCCCTCATGCAGCAGGAGGACAAGGAAGACGAACGTCTGCCG

841 ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaAlaGluAlaAlaLeuGluAsn
 GCGTCTGCTCCTGCTGTGGATGATGCTACTCATATCCCAAGCGGAAGCGGCTTGAGA
 CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTGCCCTCGCCGAAACCTCT

901 LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal
 ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGACGCACGGTCTTGTATCCTTCCTCG
 TGGAGCATTATGAATTACGTCGTAGGGACCAGGCCCTGCGTGCCAGAACATAGGAAGGAGC

961 PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe
 TGTTCTCTGCTTGCATGGTATCTGAAGGGTAAGTGGGTGCCGGAGCGGTCTACACCT
 ACAAGAACGAAACGTACCATAGACTCCATTACCCACGGGCCTGCCAGATGTGGA

1021 TyrGlyMetTrpProLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu
 TCTACGGGATGTGGCCTCTCCTGCTCTGTTGGCGTTGCCGGAGCGGGTACGCC
 AGATGCCCTACACCGGAGAGGGAGGACGAGGACAACCGAACGGGGTCGCCATGCCG

1081 AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr
 TGGACACGGAGGTGGCCCGCGTCGTGGCGGTGTTCTGTCGGGTTGATGGCGCTAA
 ACCTGTGCCTCCACCGCGCACACCGCCACAAGAGCAGCCAACACCGCATTACCGCGATT

1141 LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu
 CTCTGTCACCATAATTACAAGCGCTATATCAGCTGGTGTGGCTTCAGTATTTC
 GAGACAGTGGTATAATGTTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAG

1201 ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg
 TGACCAGAGTGGAAAGCGCAACTGCACGIGTGGATTCCCCCCCCTCAACGTCCGAGGGGGC
 ACTGGTCTCACCTCGCGTTGACGTGCACACCTAAGGGGGAGTTGCAGGCTCCCCCG

1261 AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys
 GCGACGCTGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGGTATTGACATCACCA
 CGCTGCGACAGTAGAACATGAGTACACACGACATGTGGGCTGAGACCATAAAACTGTAGTGGT

1321 LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAla
 AATGCTGCTGGCGTCTCGGACCCCTTGATTCTCAAGCCAG
 TTAACGACGACCGGAGAACCTGGGGAAACCTAACGAGTICGGTC

FIG. 46-2

FIG. 47-1 COMBINED CRF CF DNAs K9-1 through 15e

GlyCysProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGly
 1 CAGGCTGCTTGAGAGGGCTAGCCAGCTGCCGACCCCTTACCGATTTGACCAGGGCTGGG
 GTCCGACAGGACTCTCGATCGGTGACGGCTGGGAATGGCTAAACTGGTCCCACCC

 ProIleSerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrPro
 61 GCCCTATCAGTTATGCCAACGGAGCGGCCCTACTGCTGGCACTTAC
 CGGGATAGTCATAACGGTTGCCCTCGCCGGGCTGGTCGCGGGATGACGACCGTGATGG

 ProLysProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThr
 121 CCCCAAACCTGCGGTATTGTGCCCGAAGAGTGTGTGGTCCGGTATATTGCTTCA
 GGGGTTTGGAACGCCATAACACGGGCGCTTCTCACACACACCAGGCCATAACGAAGT

 ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly
 181 CTCCCAGCCCCGTGGTGGGAACGACCGACAGGTGGCGGCCACCTACAGCTGGG
 GAGGGTGGGGCACCACCCACCTGTGGCTGTCCAGCCCCGCGGGTGGATGTCGACCC

 GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe
 241 GTGAAAATGATACGGACGTCTCGTCTTAACAATACCAGGCCACCGCTGGCAATTGGT
 CACTTTACTATGCCTGAGAACAGGAATTGTTATGGTCCGGTGGCAGCCGTTAACCA

 GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal
 301 TCGGTTGACCTGGATGAACTCAACTGGATTCAACAAAGTGTGGAGGCCCTCCTTG
 AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTCACACGCCCTCGGGAGGAACAC

 IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro
 361 TCATCGGAGGGGGGGCAACACACCCCTGCACTGCCCACTGATTGCTCCGCAAGCATC
 AGTAGCCTCCCCGCCGTGTGTGGACGTGACGGGGTACTAACGAAGGCCGTTCTGAG

 AspAlaThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAsp
 421 CGGACGCCACATACTCTGGTGCGGCTCCGGTCCGGATCACACCCAGGIGCCTGGTCG
 GCCTGGGGTGTATGAGAGGCCACGCCAGGCCAGGGACCTAGTGTGGTCCACGGACCAAGC

 TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg
 481 ACTACCCGTATAGGCTTGGCATTATCCTTGTACCATCAACTACACCATAATTAAATCA
 TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAAATTAGT

 MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu
 541 GGATGTACGTGGGAGGGGTCGAACACAGGCTGGAAAGCTGCCCTGCAACTGGACGCCGGCG
 CCTACATGCACCCCTCCCCAGCTGTGTCCGACCTCGACGGACGTGACCTGCGCCCCGC

 ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr
 601 AACGTTGCGATCTGGAAAGACAGGGACAGGTCCGAGCTCAGCCGTTACTGCTGACCACTA
 TTGCAACGCTAGACCTCTGTCCCTGTCCAGGCTCGAGTCGGCAATGACGACTGGTG

 GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIle
 661 CACAGTGGCAGGTCCCTCCGTGTTCTCACACCCCTACCAGGCCITGTCCACCGGCCCTCA
 GTGTCACCGTCCAGGAGGGACAAGGAAGTGTGGATGGTCCGAACAGGTGGCCGGAGT

 HisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAla
 721 TCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGTCAAGCATCG
 AGGTGGAGGTGGTCTGTAAACACCTGCACTGACGTATGAAACATGCCCAACCCAGTTCTGAG

 SerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArg
 781 CGTCCTGGGCCATTAAAGTGGGAGTACGTGTTCTCTGTGCTTGTGAGACGCCGCG
 GCAGGACCCGGTAATTCACCCCTATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCG

 ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaLeuGluAsn
 841 GCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAGGCCGCTTGAGA
 CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTGCCCTCCGCCGAAACCTCT

 LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal
 901 ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGACGCACGGCTTGTATCCTTCTCG
 TGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTCGTGCAGAACATAGGAAGGAGC

961 PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe
 TGTTCTTCTGCTTGATGGTATTGAAGGGTAAGTGGGTGCCCGAGCGGTCTACACCT
 ACAAGAACGAAACGTACCATAAATCCCATTACCCACGGGCCTCGCAGATGTGGA

1021 TyrGlyMetTrpProLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu
 TCTACGGGATGTGGCCTCTCCTCGTGTGGCGTTGCCAGCGGGTACGCGC
 AGATGCCCTACACCGGAGAGGAGCAGAGAACCGCAACGGGTGCCCATGCGCG

1081 AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr
 TGGACACGGAGGTGGCCGCGTCGTGTGGCGTGTGGCTCGTGGGTTGATGGCGCTGA
 ACCTGTGCCCTCACCGCGCAGCACACGCCACAACAAGAGCAGCCCACCGCGACT

1141 LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu
 CTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGTGGCTCAGTATTTTC
 GAGACAGTGGTATAATGTCGCGATATAGTCGACCACGAACACCACCGAAGTCATAAAG

1201 ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg
 TGACCAGAGTGGAAAGCGCAACTGCACGTGTGGATTCCCCCTCAACGTCCGAGGGGGC
 ACTGGTCTCACCTCGCGTGTGACGTGACACCTAACGGGGAGTTGCAGGCTCCCCCG

1261 AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys
 GCGACGCCGTACCTACTCATGTGTGCTGTACACCGACTCTGGTATTGACATCACCA
 CGCTGCGCAGTAGAACGATCACGACATGTGGGCTGAGAACATAACTGTAGTGGT

1321 LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValPro
 AATTGCTGCTGGCCGTCTCGGACCCCTTGGATTCTCAAGCCAGTTGCTTAAAGTAC
 TTAAACGACGACCGGAGAACCTAACGAGTTGGGAAAGACGCGCAATCGGCCCTACTAGC

1381 TyrPheValArgValGlnGlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGly
 CCTACTTGTGCGCGTCCAAGGCCTCTCGGTTAGCGCGGAAGATGATCG
 GGATGAAACACGCGCAGGTTCCCGAAGAGGCAAGACGCGCAATCGGCCCTACTAGC

1441 GlyHisTyrValGlnMetValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyr
 GAGGCCATTACGTCAAATGGTCATCATTAAGTTAGGGCGCTACTGGCACCTATGTTT
 CTCCGGTAATGCACGTTACCACTAGTAATTCAATCCCCCGAATGACCGTGGATACAAA

1501 AsnHisLeuThrProLeuArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAla
 ATAACCATCTCACTCCTCTCGGGACTGGCGCACACGGCTTGCAGATCTGGCGTGG
 TATTGGTAGAGTGAGGAGAACCCGTGACCGCTGTGCCAACGCTCTAGACCGGACCC

1561 ValGluProValValPheSerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThr
 CTGTAGAGCCAGTCGTCTCTCCCAAATGGAGACCAAGCTCATACGTGGGGGGCAGATA
 GACATCTCGGTACGAGAACGGGTTACCTCTGGTGTGAGTAGTGCACCCCCGTCTAT

1621 AlaAlaCysGlyAspIleIleAsnGlyLeuProValSerAlaArgArgGlyArgGluIle
 CCGCCGCGTGCCTGACATCATCAACGGCTTGCCTTTCCGCCGAGGGGCCGGAGA
 GGCAGCGCACGCCACTGTAGTAGTTGCCAACGGACAAAGCGGGCTCCCCGGCCCTCT

1681 LeuLeuGlyProAlaAspGlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThr
 TACTGCTGGGCCAGCCGATGGAATGGCTCCAAGGGGTGGAGGTTGCTGGCGCCCACATCA
 ATGACGAGCCCCGGTGGCTACCTTACCAAGAGGTTCCCACCTCAACGACCGCGGGTAGT

1741 AlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArg
 CGGCGTACGCCAGCAGACAAGGGGCCTCCTAGGGTGCATAATACCAGCTAACTGGCC
 GCCGCATCGGGTCTGTTCCCGAGGATCCCACGTATTAGTGGTGGATTGACCGG

1801 AspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeu
 GGGACAAAAACCAAGTGGAGGGTGGAGGTCCAGATTGTCACACTGCTGCCAACCTTCC
 CCCTGTTTGGTTACCTCCACTCCAGGTCTAACACAGTTGACGACGGGTTGGAAAGG

1861 AlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIle
 TGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACACGGGCCGGAACGAGGACCA
 ACCGTTGCACGTAGTTACCCACACGACACTGACAGATGGTGGCCCCGGCTTGCTCCTGGT

AlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspGlnAspLeuValGly

1921 TCGCGTCACCCAAGGGTCCCTGTCATCCAGATGTATACCAATGTAGAACCAAGAACCTTG
 AGCGCAGTGGGTCCCCAGGACAGTAGGTCTACATATGGTTACATCTGGTTCTGGAACACC
 TrpProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu
 1981 GCTGGCCCGCTCCGCAAGGTAGCCGCTATTGACACCCCTGCACCTGCGGCTCCCTCGGACC
 CGACCGGGCGAGGCCGTTCCATCGCGAGTAAGTGGGACGTGAACGCCGAGGAGCCTGG
 TyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGly
 2041 TTTACCTGGTCACGAGGCACGCCGATGTCATTCCGTGCGCCGGCGGGGTGATAGCAGGG
 AAATGGACCATGCTCCGTGCGGCTACAGTAAGGCCACGCCGCCCCACTATCGTCCC
 SerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeu
 2101 GCAGCCTGCTGTCGCCCGGCCATTCTACTTGAAAGGCTCTCGGGGGTCCGCTGT
 CGTCGGACGACAGCGGGGCCGGTAAAGGATGAACCTTCCGAGGGAGCCCCCAGCGACA
 CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAla
 2161 TGTCGCCCGGGGACGCCGTGGCATATTAGGGCCGCGGTGTCACCCGTGGAGTGG
 ACACGGGGCGCCCGTGCACCGTATAAATCCCGCGCACACGTGGCACCTCAC
 LysAlaValAspPheIleProValGluAsnLeuGluThrThrMetArgSerProValPhe
 2221 CTAAGGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAAACATGAGGTCCCCGGTGT
 GATTCCGCCACCTGAAATAGGGACACCTCTGGATCTCTGTTGGTACTCCAGGGGCCACA
 ThrAspAsnSerSerProProValValProGlnSerPheGlnValAlaHisLeuHisAla
 2281 TCACGGATAACTCCTCTCCACCAGTAGTGCCCCAGAGCTTCCAGGTGGCTCACCTCCATG
 AGTGCCTATTGAGGAGAGGTGGTACACGGGTCTCGAAGGTCCACCGAGTGGAGGTAC
 ProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLys
 2341 CTCCCACAGGCAGCGGCAAAAGCACCAAGGTCCCGCTGCATATGCAGCTCAGGGCTATA
 GAGGGITGCCGTGCCGTTCTGGTTCCAGGGCCACGTATACTCGAGTCCCAGTAT
 ValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLys
 2401 AGGTGCTAGTACTCAACCCCTCTGGCTGCAACACTGGGCTTGGTGTACATGTC
 TCCACGATCATGAGTTGGGAGACAACGACGTGTAACCGAAACACGAATGTACAGGT
 AlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIleThrThrGlySerPro
 2461 AGGCTCATGGGATCGATCCTAACATCAGGACCGGGGTGAGAACATTACCACTGGCAGCC
 TCCGAGTAGCCTAGCTAGGATTGTAGTCCTGGCCCCACTCTGTTAATGGTACCGTC
 IleThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyr
 2521 CCATCACGTACTCACCTACGGCAAGTCCCTIGCCGACGGCGGGTGCTCGGGGGCGCTT
 GGTAGTGCATGAGGTGGATGCCGTTCAAGGAACCGCTGCCGCCACGAGCCCCCGCGAA
 AspIleIleIleCysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGly
 2581 ATGACATAATAATTGTGACGAGTGCACACTCCACGGATGCCACATCCCATCTGGGCATCG
 TACTGTATTATAAACACTGCTCACGGTGAGGTGCCTACGGTAGGTAGAACCCGTAGC
 ThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValLeuAlaThrAlaThr
 2641 GCACTGTCCTGACCAAGCAGAGACTGGGGGGCGAGACTGGGTGTGCTGCCACCGCCA
 CGTGACAGGAACCTGGTCTGACGCCCGCTCTGACCAACACGAGCGGTGGCGGT
 ProProGlySerValThrValProHisProAsnIleGluGluValAlaLeuSerThrThr
 2701 CCCCTCCGGGCTCCGTACTGTGCCCATCCAAACATCGAGGGAGTTGCTCTGTC
 GGGGAGGGCCGAGGCAGTGACACGGGGTAGGGTTGAGTCCTCCAAACGAGACAGGTGGT
 GlyGluIleProPheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHis
 2761 CCGGAGAGATCCCCTTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGGAGAC
 GGCCTCTAGGGAAAAATGCCGTTCCGATAGGGGAGCTTCATTAGTCCCCCCCCTCTG
 LeuIlePheCysHisSerLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeu
 2821 ATCTCATCTCTGTCAATTCAAAGAAGAAGTGCACGAACTCGCCGAAAGCTGGTC
 TAGAGTAGAACAGTAAGTTCTTCACGCTGCTTGAGCGCGTTCGACCAGCGTA
 GlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGly
 2881 TGGGCATCAATGCCGTGGCTACTACCGCGGTCTGACGTGTCATCCGACCAGCG
 ACCCGTAGTTACGGCACCGGATGATGGCGCCAGAACACTGCACAGGCAGTAGGGCTGGTC

2941 AspValValValValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSer
 GCGATTTGTCGTCGTGGCAACCGATGCCCTCATGACCGGCTATACCGCGACTTCGACT
 CGCTACAAACAGCAGCACCGTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGA

3001 ValIleAspCysAsnThrCysValThrValAspPheSerLeuAspProThrPhe
 CGGTGATAGACTGCAATACTGTTGTCACCCAGACAGTCGATTTCAGCCTTGACCTACCT
 GCCACTATCTGACGTTATGCACACAGTGGGTCTGTCAGCTAAAGTCGGAACTGGGATGGA

3061 ThrIleGluThrIleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArg
 TCACCATTGAGACAATCACGCTCCCCCAGGATGCTGCTCCCGACTCAACGTCGGGGCA
 AGTGGTAACTCTGTTAGTGCAGGGGGTCTACGACAGAGGGCGTGAGTTGCAGCCCCGT

3121 ThrGlyArgGlyLysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGly
 GGACTGGCAGGGGGAAAGCCAGGCATCTACAGATTGTCGGCACCGGGGGAGCGCCCTCCG
 CCTGACCGTCCCCCTCGGTCCGTAGATGTCTAACACCGTGGCCCCCTCGCGGGGAGGC

3181 MetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeu
 GCATGTTGACTCGTCGCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGC
 CGTACAAGCTGAGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCG

3241 ThrProAlaGluThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProVal
 TCACGCCCGCCGAGACTACAGTTAGGCTACGAGCGTACATGAAACACCCGGGGCTCCCG
 AGTGCAGGGCGGCTCTGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGC

3301 CysGlnAspHisLeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAla
 TGTGCCAGGACCATCTGAATTGGGAGGGCGCTTACAGGCCTACTCATATAGATG
 ACACGGTCTGGTAGAACTTAAACCTCCCGCAGAAATGTCCGGAGTGAATATCTAC

3361 HisPheLeuSerGlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGln
 CCCACTTCTATCCCAGACAAAGCAGAGTGGGGAGAACCTTCCCTACCTGGTAGCGTACC
 GGGTAGAAAGATAGGGCTGTTGCTCACCCCTCTTGGAAAGGAATGGACCATCGCATGG

3421 AlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCys
 AAGCCACCGTGCGCTAGGGCTCAAGCCCCTCCCCATGTCGGGACAGATGTGGAAAGT
 TTGGTGGCACACCGCATCCGAGTTGCGGGAGGGTAGCACCTGGTCTACACCTCA

3481 LeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAla
 GTTGATTCGCTCAAGCCCACCCCTCCATGGGCCAACACCCCTGCTATACAGACTGGCG
 CAAACTAAGCGGAGTTGGGGAGGTACCCGGTTGTGGGACGATAATGCTGACCCGC

3541 ValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSer
 CTGTCAGAACATGAAATCACCTGACGCACCCAGTCACAAATACATCATGACATGCATGT
 GACAAGTCTTACTTTAGTGGGACTGCGTGGGTCACTGGTTATGTAGTACTGTACGTACA

3601 AlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeu
 CGGCCGACCTGGAGGTCGTCACGAGCACCTGGGTGCTCGTGGCGCGTCTGGCTGCT
 GCCGGCTGGACCTCCAGCAGTGTGCGACCCACGAGCAACCGCCGAGGACCGACGAA

3661 AlaAlaTyrCysLeuSerThrGlyCysValValIleValGlyArgValValLeuSerGly
 TGGCCGCGTATTGCTGTCAACAGGCTGCGTGGTCATAGTGGCAGGGTCGTCTGTCCG
 ACCGGCGCATAACGGACAGTTGTCGACCGACCAGTATCACCCGCTCCAGCAGAACAGGC

3721 LysProAlaIleIleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGlu
 GGAAGCCGGCAATCACCTGACAGGGAGTCCTCTACCGAGAGGTTGATGAGATGGAAG
 CCTTCGGCGTTAGTATGGACTGTCCTCAGGAGATGGCTCTCAAGCTACTCTACCTTC

3781 CysSerGlnHisLeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGln
 AGTGCTCTCAGCACCTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGC
 TCACGAGAGTCGTGAATGGCATGTAGCTGTTCCCTACTACGAGCGGCTCGTCAAGTCG

3841 LysAlaLeuGlyLeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaVal
 AGAAGGCCCTCGGCCCTCGAGACCGCGTCCCGTACGGCAGAGGTTATCGCCCTGCTG
 TCTTCGGGAGCCGGAGGACGTCTGGCGCAGGGCAGTCGTCTCCAATAGCGGGGACGAC

GlnThrAsnTrpGlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSer

3901 TCCAGACCAACTGGCAAAACTCGAGACCTCTGGCGAAGCATATGTGGAACATCATCA
 AGGTCTGGT1GACCCTTIGAGCTCTGGAAGACCCGCTTCGTATAACACCTTGAAAGTAGT
 GlyIleGlnTyrLeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeu
 3961 GTGGGATACAATACTGGCGGGCTTGTCAACGCTGCCCTGGTAACCCGCCATTGCTTCAT
 CACCCCTATGTTATGAAACGCCGAACAGTTGCGACGGACCATTGGGCGGTAAACGAAAGTA
 MetAlaPheThrAlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsn
 4021 TGATGGCTTTACAGCTGCTGTCACCAGCCCCTAACCACTAGCCAACCCCTCCTCTCA
 ACTACCGAAAATGTCGACGACAGTGGTCGGGTGATTGGTGATGGTTGGGAGGAGAAGT
 IleLeuGlyGlyTrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheVal
 4081 ACATATTGGGGGGGGTGGGTGGCTGCCAGCTGCCGCCGGTGCCGCTACTGCCTTG
 TGTATAACCCCCCACCACCGACGGTCGAGCGGGGGGCCACGGCGATGACGGAAAC
 GlyAlaGlyLeuAlaAlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAsp
 4141 TGGGCGCTGGCTTAGCTGGCGCCCATCGGCAGTGGACTGGGAAGGTCCCTCATAG
 ACCCGCGACCGAATCGACCGCGCGTAGCCGTACAAACCTGACCCCTCCAGGAGTATC
 IleLeuAlaGlyTyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSer
 4201 ACATCCTTGCAAGGGTATGGCGCGGCGTGGCGGGAGCTTGTGGCATTCAAGATCATGA
 TGTAGGAACGTCCCATAACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTAGTACT
 GlyGluValProSerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGly
 4261 GCGGTGAGGTCCCCTCCACGGAGGACCTGGTCAATCTACTGCCGCCATCCTCTGCCCG
 CGCCACTCCAGGGAGGTGCCTCTGGACCAGTTAGATGACGGCGGTAGGAGAGCGGGC
 AlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGlu
 4321 GAGCCCTCGTAGTCGGCGTGGCTGTGCAGCAATACTGCGCCGCACGGTGGCCGGCG
 CTCGGGAGCATCAGCCGACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCG
 GlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSer
 4381 AGGGGGCAGTGCAGTGGATGAAACGGCTGATAGCCTCGCCTCCGGGGAAACCATGTTT
 TCCCCCGTCACGTACCTACTTGGCCGACTATCGGAAGCGGGCCCTTGGTACAAA
 ProThrHisTyrValProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSer
 4441 CCCCCACGCACTACGTGCCGGAGAGCGATGCAGCTGCCCGCTACTGCCATACTCAGCA
 GGGGGTGCCTGATGCACGGCCTCTCGCTACGTCGACGGCGCAGTGACGGTATGAGTCG
 LeuThrValThrGlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThr
 4501 GCCTCACTGTAACCCAGCTCCTGAGGCAGTGCACCACTGGATAAGCTGGAGTGTACCA
 CGGAGTGACATTGGGTCGAGGACTCCGCTGACGTGGTACCTATTGAGCCTCACATGGT
 ProCysSerGlySerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAsp
 4561 CTCCATGCTCCGGTCCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGAGCG
 GAGGTACGAGGCCAAGGACCGATTCCCTGTAGACCTATACGCTCCACAACCTCG
 PheLysThrTrpLeuLysAlaLysLeuMetProGlnLeuProGlyIleProPheValSer
 4621 ACTTTAACGACTGGCTAAAAGCTAACGCTCATGCCACAGCTGCCCTGGATCCCCTTG
 TGAAATTCTGGACCGATTTCGATTGAGTACGGTGTGACGGACCCCTAGGGAAACACA
 CysGlnArgGlyTyrLysGlyValTrpArgValAspGlyIleMetHisThrArgCysHis
 4681 CCTGCCAGCGCGGGTATAAGGGGGTCTGGCAGTGGACGGCATATGCACACTCGCTGCC
 GGACGGTCGCGCCCATATTCCCCCAGACCGCTACCTGCCGTAGTA~~G~~TGAGCGACGG
 CysGlyAlaGluIleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArg
 4741 ACTGTGGAGCTGAGATCACTGGACATGTC~~CAAA~~ACGGGACGATGAGGATGTCGGTCCTA
 TGACACCTCGACTCTAGTGACCTGTACAGTTTGCCTGCTACTCCTAGCAGCCAGGAT
 ThrCysArgAsnMetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCys
 4801 GGACCTGAGGAACATGTGGAGTGGACCTCCCCATTAAATGCCTACACCACGGGCC
 CCTGGACGTCCTGTACACCTCACCCCTGGAAAGGGTAATTACGGATGTGGTGC
 CATGGGGGAAGGACGCCGTTGATGTGCAAGCGCGATAACCTCCCACAGACGTCTCCTTA

4921 ValGluIleArgGlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeu
 ATGTGGAGATAAGGCAGGTGGGGACTTCACTACGTGACGGGTATGACTACTGACAATC
 TACACCTCTATTCCGTCCACCCCTGAAGGTGATGCACTGCCATACTGATGACTGTTAG

4981 LysCysProCysGlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeu
 TCAAATGCCGTGCCAGGTCCCCTCGCCCCAATTTACAGAATTGGACGGGTGCGCC
 AGTTTACGGGCACGGTCCAGGGTAGCGGGCTAAAAAGTGTCTAACCTGCCACGCGG

5041 HisArgPheAlaProProCysLysProLeuLeuArgGluGluValSerPheArgValGly
 TACATAGGTTGCGCCCCCTGCAAGCCCTTGCTGCGGGAGGGTATCATTAGAGTAG
 ATGTATCCAAACGCCGGGGACGTTCGGGAACGACGCCCTCCATAGTAAGTCTCATC

5101 LeuHisGluTyrProValGlySerGlnLeuProCysGluProGluProAspValAlaVal
 GACTCCACGAATAACCGGTAGGGTCGCAATTACCTGCGAGCCGAACCGGACGTGGCCG
 CTGAGGTGCTTATGGCCATCCCAGCGTAATGGAACGCTCGGGCTTGGCTGCACCGGC

5161 LeuThrSerMetLeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeu
 TGTTGACGTCCATGCTCACTGATCCCTCCATATAACAGCAGAGGCGGGCGAAGGT
 ACAACTGCAGGTACGAGTGACTAGGGAGGGTATATTGTCGTCTCCGCCGGCGCTCCA

5221 AlaArgGlySerProProSerValAlaSerSerAlaSerGlnLeuSerAlaProSer
 TGGCGAGGGATCACCCCCCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCAT
 ACCGCTCCCTAGTGGGGGAGACACCGGTCGAGGAGCCGATCGGTGATAGGCGAGGTA

5281 LeuLysAlaThrCysThrAlaAsnHisAspSerProAspAlaGluIleGluAlaAsn
 CTCTCAAGGCAACTTGACCGCTAACCATGACTCCCTGATGCTGAGCTCATAGAGGCCA
 GAGAGTTCCGTTGAACGTGGCATTGGTACTGAGGGACTACGACTCGAGTATCTCCGGT

5341 LeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysVal
 ACCTCCTATGGAGGCAGGAGATGGGCGAACATCACCAGGGTTGAGTCAGAAAACAAAG
 TGGAGGATACTCCGCTCTACCCGCCGTTGAGTGGTCCAACTCAGTCTTTGTTTC

5401 ValIleLeuAspSerPheAspProLeuValAlaGluGluAspGluArgGluIleSerVal
 TGGTGATCTGGACTCCTCGATCCGCTTGTGGCGAGGAGACGAGCGGGAGATCTCCG
 ACCACTAAGACCTGAGGAAGCTAGGCGAACACCGCCTCCTGCTGCCCTAGAGGC

5461 ProAlaGluIleLeuArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArg
 TACCCGCAGAAATCTCGCGGAAGTCTCGGAGATTCCGCCAGGCCCTGCCGTTGGCGC
 ATGGCGCTTTAGGACGCCCTCAGAGCCTTAAGCGGGTCCGGACGGCAAACCCGCG

5521 ProAspTyrAsnProProLeuValGluThrTrpLysProAspTyrGluProProVal
 GGCGGACTATAACCCCCCGCTAGTGGAGACGTGGAAAAGCCGACTACGAACCACCTG
 CGGCCCTGATATTGGGGGGCGATCACCTCTGCACCTTTCGGGCTGATGCTTGGTGGAC

5581 ValHisGlyCysProLeuProProLysSerProProValProProArgLysLys
 TGGTCCATGGCTCCGCTTCCACCTCCAAAGTCCCTCTGTGCCTCCGCCCTGGAAAGA
 ACCAGGTACCGACAGCGAAGGTGGAGGTTCAAGGGAGGACACGGAGGCGAGCCTCT

5641 ArgThrValValLeuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArg
 AGCGGACGGTGGTCTCACTGAATCAACCCCTATCTACTGCCTGGCCGAGCTGCCACCA
 TCGCCTGCCACCAGGAGTGACTTAGTGGGATAGATGACGGAACCGGCTCGAGCGGTGGT

5701 SerPheGlySerSerSerThrSerGlyIleThrGlyAspAsnThrThrSerSerGlu
 GAAGCTTGGCAGCTCCTCAACTTCCGGCATTACGGCGACAATACGACAACATCCTCTG
 CTTCGAAACCGTCGAGGAGTTGAAGGCCGTAATGCCGCTGTTATGCTGTAGGAGAC

5761 ProAlaProSerGlyCysProProAspSerAspAlaGluSerTyrSerSerMetProPro
 AGCCCGCCCTCTGGCTGCCCGGGCGACTCCGACGCTGAGTCCTATTCCCTCATGCC
 TCGGGCGGGGAAGACCGACGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGG

5821 LeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSer
 CCCTGGAGGGGGAGGCTGGGATCCGGATCTAGCGACGGGTATGGTCAACGGTCAGTA
 GGGACCTCCCCCTGGACCCCTAGGCCTAGAATCGCTGCCAGTACCAAGTGGCAGTCAT

GluAlaAsnAlaGluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeu

5881 GTGAGGCCAACGCGGAGGATGICGTGTGCTCAATGCTTACTCTTGGACAGGCGCAC
 CACTCCGGTTGCGCCTCCTACAGCACACGAGTTACAGAATGAGAACCTGTCCCGCTG
 ValThrProCysAlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeu
 5941 TCGTCACCCCGTGCGCCGCGGAAGAACAGAAACTGCCCATCAATGCACTAAAGCAACTCGT
 AGCAGTGGGGCACGCCGCGCTCTTGCTTACGGTAGTTACGTGATTGTTGAGCA
 LeuArgHisHisAsnLeuValTyrSerThrSerArgSerAlaCysGlnArgGlnLys
 6001 TGCTACGTCAACCACAATTGGTGTATTCCACCACCTCACGAGTGTGCAAAAGGCAGA
 ACGATGCAGTGGTGTAAACACATAAGGTGGAGTGCACGTCACGAACGGTTTCCGTCT
 LysValThrPheAspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGlu
 6061 AGAAAGTCACATTGACAGACTGCAAGTCTGGACAGCCATTACCAAGGACGTACTCAAGG
 TCTTTCAGTGTAAACTGTCACGTTCAAGACCTGTCGTAATGGTCTGCATGAGTTCC
 ValLysAlaAlaAlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSer
 6121 AGGTTAAAGCAGCGCGTCAAAGTGAAGGCTAACTTGCTATCGTAGAGGAAGCTTGCA
 TCCAATTGTCGCCGCAGTTTCACTTCGATTGAACGATAGGCATCTCTCGAACGCT
 LeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCys
 6181 GCCTGACGCCCAACACTCAGCAAATCAAAGTTIGTTATGGGGCAAAAGACGTCCGTT
 CGGACTGCGGGGGTGTGAGTCGTTAGGTCAAACCAATACCCGTTTCTGCAGGCAA
 HisAlaArgLysAlaValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsn
 6241 GCCATGCCAGAAAGGCCGTAACCCACATCAACTCCGTGAGAACACCTCTGGAAGACA
 CGGTACGGTCTTCCGGATTGGTGTAGTTGAGGCACACCTCTGGAAGACCTCTGT
 ValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGlu
 6301 ATGTAACACCAATAGACACTACCACATGGCTAAAGAACGAGGTTTCTGCCTCAGCC
 TACATTGTGGTTATCTGTGATGGTAGTACCGATTCTGCTCAAAGACGCAAGTCGGAC
 LysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGlyValArgValCys
 6361 AGAAGGGGGTCGTAAGCCAGCTCGTCTCAICGTGTCCTCCGATCTGGCGTGCCTGT
 TCTCCCCCAGCATTGGTCGAGCAGAGTAGCACAAGGGCTAGACCGCACGCGCAC
 GluLysMetAlaLeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSer
 6421 GCGAAAAGATGGCTTGACAGCTGGTACAAAGCTCCCTGGCGTGAIGGGAAAGCT
 CGCTTTCTACCGAAACATGCTGCACCAATGTTGAGGGAAACCGGACTACCTTCGA
 TyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSer
 6481 CCTACGGATTCAAATACTCACCAGGACAGCGGGTIGAAITCCTCGTGCAGCGTGGAA
 GGATGCCTAAGGTTATGAGTGGCCTGTCGCCAACTTAAGGAGCACGTTGACCTTC
 LysLysThrProMetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGlu
 6541 CCAAGAAAACCCCAATGGGTTCTCGTATGATACCCGCTGCTTGACTCCACAGTC
 GGTTCTTGGGGTACCCCAAGAGCATACTATGGCGACGAAACTGAGGTGTCAGTGAC
 SerAspIleArgThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArg
 6601 AGAGCGACATCCGTACGGAGGAGGCAACTACCAATGTTGTGACCTCGACCCCCAAGCCC
 TCTCGCTGAGGCATGCCCTCCGTTAGATGGTTACAACACTGGAGCTGGGGTTCCGG
 ValAlaIleLysSerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArg
 6661 GCGTGGCCATCAAGTCCCTACCGAGAGGCTTATGTTGGGGGCCCTCTAACCAATTCAA
 CGCACCGGTAGTTAGGGAGTGGCTCTCGAAATACAACCCCGGGAGATGGTTAAGTT
 GlyGluAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGly
 6721 GGGGGGAGAACTGCGGCTATCGCAGGTGCCGCGAGCGTACTGACAACACTAGCTGTG
 CCCCTCTTGACGCCGATAGCGTCCACGGCGCGCTGCCGCAITGACTGTTGATCGACAC
 AsnThrLeuThrCysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAsp
 6781 GTAACACCCCTCACTTGCTACATCAAGGCCGGCAGCCTGTCAGCCGCAAGGGCTCCAGG
 CATTGTGGGAGTGAACGATGTAGTTCCGGGCCGTCGGACAGCTCGCGTCCCGAGGTCC
 CysThrMetLeuValCysGlyAspAspLeuValValIleCysGluSerAlaGlyValGln
 6841 ACTGCACCATGCTCGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGGTCC
 TGACGTGGTACGAGCACACACCGCTGCTGAATCAGCAATAGACACTTCGCGCCCCCAGG

63

2212511

GluAspAlaAlaSerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProPro
6901 AGGAGGACGCCGGCGAGCCTGAGAGCCTCACGGAGGCTATGACCAGGTACTCCGCCCCC
TCCTCCTGCGCCGCTCGGACTCTCGGAAGTGCCCTCGATACTGGTCCATGAGGCGGGGGG
GlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnVal
6961 CTGGGGACCCCCCAACACAGAAATCGACTTGGAGCTATAACATCATGCTCCTCCAACG
GACCCCTGGGGGGTGTGGTCTTATGCTGAAACCTCGAGTATTGTAGTACGAGGAGGTTGC
SerValAlaHisAspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThr
7021 TGTCAGTCGCCAACGCGCGCTGGAAAGAGGGTCTACTACCTCACCGTGACCCCTACAA
ACAGTCAGCGGGTGTGCGCGACCTTCTCCAGATGATGGAGTGGGACTGGGATGTT
ProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGly
7081 CCCCCCTCGCGAGAGCTGCGTGGGAGACAGCAAGACACACTCCAGTCATTCCCTGGCTAG
GGGGGGAGCGCTCTGACGCACCCCTCTGCGTTCTGTGAGGTCAAGGACCGATC
AsnIleIleMetPheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePhe
7141 GCAACATAATCATGTTGCCAACACTGTGGCGAGGATGATACTGATGACCCATTCT
CGTTGTATTAGTACAAACGGGGGTGTGACACCCGCTCTACTATGACTACTGGGTAAAGA
SerValLeuIleAlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAla
7201 TTAGCGCTTATAGCCAGGGACCAGCTTGAAACAGGCCCTCGATTGCGAGATCTACGGGG
AATCGCAGGAATATCGTCCCTGGTCGAACTTGTCCGGGAGCTAACGCTCTAGATGCC
CysTyrSerIleGluProLeuAspLeuProProIleIleGlnArgLeu
7261 CCTGCTACTCCATAGAACCACTTGATCTACCTCCAATTCAGAACACTC
GGACGATGAGGTATCTGGTGAAGTAGATGGAGGTAGTAAGTTCTGAG

FIG. 47-8

2212511

5

NANBV DIAGNOSTICS AND VACCINES

10

15

20

Technical Field

The invention relates to materials and methodologies for managing the spread of non-A, non-B hepatitis virus (NANBV) infection. More specifically, it relates to diagnostic DNA fragments, diagnostic proteins, diagnostic antibodies and protective antigens and antibodies for an etiologic agent of NANB hepatitis, i.e., hepatitis C virus.

30

References Cited in the Application

- Barr et al. (1986), Biotechniques 4:428.
Botstein (1979), Gene 8:17.
Brinton, M.A. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND
35 FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol.

- eds. Schlesinger and Schlesinger, Plenum Press), p.327-374.
- Broach (1981) in: Molecular Biology of the Yeast *Saccharomyces*, Vol. 1, p.445, Cold Spring Harbor Press.
- 5 Broach et al. (1983), Meth. Enz. 101:307.
Chang et al. (1977), Nature 198:1056.
Chirgwin et al. (1979), Biochemistry 18:5294.
Chomczynski and Sacchi (1987), Analytical Biochemistry 162:156.
- 10 Clewell et al. (1969), Proc. Natl. Acad. Sci. USA 62:1159.
Clewell (1972), J. Bacteriol. 110:667.
Cohen (1972), Proc. Natl. Acad. Sci. USA 69:2110.
Cousens et al. (1987), gene 61:265.
De Boer et al. (1983), Proc. Natl. Acad. Sci. USA 292:128.
- 15 Dreesman et al. (1985), J. Infect. Disease 151:761.
Feinstone, S.M. and Hoofnagle, J.H. (1984), New Engl. J. Med. 311:185.
Fields & Knipe (1986), FUNDAMENTAL VIROLOGY (Raven Press, N.Y.).
- 20 Fiers et al. (1978), Nature 273:113.
Gerety, R.J. et al., in VIRAL HEPATITIS AND LIVER DISEASE (Vyas, B.N., Dienstag, J.L., and Hoofnagle, J.H., eds, Grune and Stratton, Inc., 1984) pp 23-47.
Goeddel et al. (1980), Nucleic Acids Res. 8:4057.
- 25 Graham and Van der Eb (1978), Virology 52:546.
Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961.
Grych et al. (1985), Nature 316:74.
Gubler and Hoffman (1983), Gene 25:263.
- 30 Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS.
Hess et al. (1968), J. Adv. Enzyme Reg 7:149.
Hinnen et al. (1978), Proc. Natl. Acad. Sci. 75:1929.
Hitzeman et al. (1980), J. Biol. Chem. 255:2073.
- 35 Holland et al. (1978), Biochemistry 17:4900.

- Holland (1981), J. Biol. Chem. 256: 1385.
Houghton et al. (1981), Nucleic Acids Res. 9:247
Hunyh, T.V. et al. (1985) in DNA CLONING TECHNIQUES; A
PRACTICAL APPROACH (D. Glover, Ed., IRL Press, Oxford,
5 U.K.) pp. 49-78.
Immun. Rev. (1982) 62:185.
Iwarson (1987), British Medical J. 295:946.
Kennett et al. (1980) MONOCLONAL ANTIBODIES.
Laemmli (1970), Nature 227, 680.
10 Lee et al. (1988), Science 239:1288.
Maniatis, T., et al. (1982) MOLECULAR CLONING; A
LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring
Harbor, N.Y.).
Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN
15 CELL AND MOLECULAR BIOLOGY (Academic Press, London).
Maxam et al. (1980), Methods in Enzymology 65:499.
MacNamara et al. (1984), Science 226:1325.
Messing et al. (1981), Nucleic Acids Res. 9:309.
Messing (1983), Methods in Enzymology 101:20-37.
20 METHODS IN ENZYMOLOGY (Academic Press).
Michelle et al., Int. Symposium on Viral Hepatitis.
Monath (1986) in THE VIRUSES: THE TOGAVIRIDAE AND
FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol.
eds. Schlesinger and Schlesinger, Plenum Press), p.375-
25 440.
Nagahuma et al. (1984), Anal. Biochem. 141:74.
Neurath et al. (1984), Science 224:392.
Nisonoff et al. (1981), Clin. Immunol. Immunopathol.
21:397-406.
30 Overby, L.R. (1985), Curr. Hepatol. 5:49.
Overby, L.R. (1986), Curr. Hepatol. 6:65.
Overby, L.R. (1987), Curr. Hepatol. 7:35.
Peleg (1969), Nature 221:193.
Pfefferkorn and Shapiro (1974), in COMPREHENSIVE VIROLOGY,

- Vol. 2 (Fraenkel-Conrat & Wagner, eds., Plenum, N.Y.) pp. 171-230.
- Prince, A.M. (1983), Annu. Rev. Microbiol. 37:217.
- Rice et al. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND
- 5 FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.279-328.
- Roehrig (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. 10 eds. Schlesinger and Schlesinger, Plenum Press)
- Sadler et al. (1980), Gene 8, 279.
- Saiki et al. (1986), Nature 324: 163.
- Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463.
- Schlesinger et al. (1986), J. Virol. 60:1153.
- 15 Schreier, M., et al. (1980) HYBRIDOMA TECHNIQUES Scopes (1984), PROTEIN PURIFICATION, PRINCIPLES AND PRACTICE, SECOND EDITION (Springer-Verlag, N.Y.).
- Shimatake et al. (1981), Nature 292:128.
- Steimer et al. (1986), J. Virol. 58:9.
- 20 Stollar (1980), in THE TOGAVIRUSES (R.W. Schlesinger, ed., Academic Press, N.Y.), pp. 584-622.
- Taylor et al. (1976), Biochem. Biophys. Acta 442:324.
- Towbin et al. (1979), Proc. Natl. Acad. Sci. USA 76, 4350.
- Tsu and Herzenberg (1980), in SELECTED METHODS IN CELLULAR
- 25 IMMUNOLOGY (W.H. Freeman and Co.) pp. 373-391.
- Vytdehaag et al. (1985), J. Immunol. 134:1225.
- Valenzuela, P., et al. (1982), Nature 298:344.
- Valenzuela, P., et al. (1984), in HEPATITIS B (Millman, I., et al., ed, Plenum Press) pp. 225-236.
- 30 Warner (1984), DNA 3:401.
- Wu and Grossman (1987), Methods in Enzymology Vol. 154, RECOMBINANT DNA, Part E.
- Wu (1987), Methods in Enzymology vol 155, RECOMBINANT DNA, part F.
- 35 Zoller (1982), Nucleic Acids Res. 10:6487.

Cited Patents

- U.S. Patent No. 4,341,761
U.S. Patent No. 4,399,121
U.S. Patent No. 4,427,783
5 U.S. Patent No. 4,444,887
U.S. Patent No. 4,466,917
U.S. Patent No. 4,472,500
U.S. Patent No. 4,491,632
U.S. Patent No. 4,493,890

10

Background Art

Non-A, Non-B hepatitis (NANBH) is a transmissible disease or family of diseases that are believed to be viral-induced, and that are distinguishable 15 from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses, i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus 20 (EBV). NANBH was first identified in transfused individuals. Transmission from man to chimpanzee and serial passage in chimpanzees provided evidence that NANBH is due to a transmissible infectious agent or agents. However, the transmissible agent responsible for NANBH is 25 still unidentified and the number of agents which are causative of the disease are unknown.

Epidemiologic evidence is suggestive that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the 30 sporadically occurring (community acquired) type. However, the number of agents which may be the causative of NANBH are unknown.

Clinical diagnosis and identification of NANBH has been accomplished primarily by exclusion of other 35 viral markers. Among the methods used to detect putative

NANBV antigens and antibodies are agar-gel diffusion, counterimmunolectrophoresis, immunofluorescence microscopy, immune electron microscopy, radioimmunoassay, and enzyme-linked immunosorbent assay. However, none of 5 these assays has proved to be sufficiently sensitive, specific, and reproducible to be used as a diagnostic test for NANBH.

Until now there has been neither clarity nor agreement as to the identity or specificity of the antigen 10 antibody systems associated with agents of NANBH. This is due, at least in part, to the prior or co-infection of HBV with NANBV in individuals, and to the known complexity of the soluble and particulate antigens associated with HBV, as well as to the integration of HBV DNA into the genome 15 of liver cells. In addition, there is the possibility that NANBH is caused by more than one infectious agent, as well as the possibility that NANBH has been misdiagnosed. Moreover, it is unclear what the serological assays detect in the serum of patients with NANBH. It has been postu- 20 lated that the agar-gel diffusion and counterimmuno- electrophoresis assays detect autoimmune responses or non- specific protein interactions that sometimes occur between serum specimens, and that they do not represent specific NANBV antigen-antibody reactions. The immunofluorescence, 25 and enzyme-linked immunosorbent, and radioimmunoassays appear to detect low levels of a rheumatoid-factor-like material that is frequently present in the serum of patients with NANBH as well as in patients with other hepatic and nonhepatic diseases. Some of the reactivity 30 detected may represent antibody to host-determined cytoplasmic antigens.

There are a number of candidate NANBV. See, for example the reviews by Prince (1983), Feinstone and Hoofnagle (1984), and Overby (1985, 1986, 1987) and the 35 article by Iwarson (1987). However, there is no proof

that any of these candidates represent the etiological agent of NANBH.

The demand for sensitive, specific methods for screening and identifying carriers of NANBV and NANBV 5 contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and NANBH accounts for up to 90% of these cases. The major problem in this disease is the frequent progression to chronic liver damage (25-55%).

10 Patient care as well as the prevention of transmission of NANBH by blood and blood products or by close personal contact require reliable diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to NANBV. In addition, there is also a 15 need for effective vaccines and immunotherapeutic therapeutic agents for the prevention and/or treatment of the disease.

Disclosure of the Invention

20 The invention pertains to the isolation and characterization of a newly discovered etiologic agent of NANBH, hepatitis C virus (HCV). More specifically, the invention provides a family of cDNA replicas of portions of HCV genome. These cDNA replicas were isolated by a 25 technique which included a novel step of screening expression products from cDNA libraries created from a particulate agent in infected tissue with sera from patients with NANBH to detect newly synthesized antigens derived from the genome of the heretofore unisolated and 30 uncharacterized viral agent, and of selecting clones which produced products which reacted immunologically only with sera from infected individuals as compared to non-infected individuals.

Studies of the nature of the genome of the HCV, 35 utilizing probes derived from the HCV cDNA, as well as

sequence information contained within the HCV cDNA, are suggestive that HCV is a Flavivirus or a Flavi-like virus.

Portions of the cDNA sequences derived from HCV are useful as probes to diagnose the presence of virus in samples, and to isolate naturally occurring variants of the virus. These cDNAs also make available polypeptide sequences of HCV antigens encoded within the HCV genome(s) and permits the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or 10 as components of vaccines. Antibodies, both polyclonal and monoclonal, directed against HCV epitopes contained within these polypeptide sequences are also useful for diagnostic tests, as therapeutic agents, for screening of antiviral agents, and for the isolation of the NANBV agent 15 from which these cDNAs derive. In addition, by utilizing probes derived from these cDNAs it is possible to isolate and sequence other portions of the HCV genome, thus giving rise to additional probes and polypeptides which are useful in the diagnosis and/or treatment, both 20 prophylactic and therapeutic, of NANBH.

Accordingly with respect to polynucleotides, some aspects of the invention are: a purified HCV polynucleotide; a recombinant HCV polynucleotide; a recombinant polynucleotide comprising a sequence derived 25 from an HCV genome or from HCV cDNA; a recombinant polynucleotide encoding an epitope of HCV; a recombinant vector containing the any of the above recombinant polynucleotides, and a host cell transformed with any of these vectors.

Other aspects of the invention are: a recombinant expression system comprising an open reading frame (ORF) of DNA derived from an HCV genome or from HCV cDNA, wherein the ORF is operably linked to a control sequence compatible with a desired host, a cell

transformed with the recombinant expression system, and a polypeptide produced by the transformed cell.

Still other aspects of the invention are:
purified HCV, a preparation of polypeptides from the
5 purified HCV; a purified HCV polypeptide; a purified
polypeptide comprising an epitope which is immunologically
identifiable with an epitope contained in HCV.

Included aspects of the invention are a recombinant HCV polypeptide; a recombinant polypeptide
10 comprised of a sequence derived from an HCV genome or from HCV cDNA; a recombinant polypeptide comprised of an HCV epitope; and a fusion polypeptide comprised of an HCV polypeptide.

Also included in the invention are a monoclonal
15 antibody directed against an HCV epitope; and a purified preparation of polyclonal antibodies directed against an HCV epitope.

Another aspect of the invention is a particle which is immunogenic against HCV infection comprising a
20 non-HCV polypeptide having an amino acid sequence capable of forming a particle when said sequence is produced in a eukaryotic host, and an HCV epitope.

Still another aspect of the invention is a polynucleotide probe for HCV.

25 Aspects of the invention which pertain to kits are those for: analyzing samples for the presence of polynucleotides derived from HCV comprising a polynucleotide probe containing a nucleotide sequence from HCV of about 8 or more nucleotides, in a suitable
30 container; analyzing samples for the presence of an HCV antigen comprising an antibody directed against the HCV antigen to be detected, in a suitable container; analyzing samples for the presence of an antibodies directed against an HCV antigen comprising a polypeptide containing an HCV

epitope present in the HCV antigen, in a suitable container.

Other aspects of the invention are: a polypeptide comprised of an HCV epitope, attached to a 5 solid substrate; and an antibody to an HCV epitope, attached to a solid substrate.

Still other aspects of the invention are: a method for producing a polypeptide containing an HCV epitope comprising incubating host cells transformed with 10 an expression vector containing a sequence encoding a polypeptide containing an HCV epitope under conditions which allow expression of said polypeptide; and a polypeptide containing an HCV epitope produced by this method.

15 The invention also includes a method for detecting HCV nucleic acids in a sample comprising reacting nucleic acids of the sample with a probe for an HCV polynucleotide under conditions which allow the formation of a polynucleotide duplex between the probe and 20 the HCV nucleic acid from the sample; and detecting a polynucleotide duplex which contains the probe.

Immunoassays are also included in the invention. These include an immunoassay for detecting an HCV antigen comprising incubating a sample suspected of containing an 25 HCV antigen with a probe antibody directed against the HCV antigen to be detected under conditions which allow the formation of an antigen-antibody complex; and detecting an antigen-antibody complex containing the probe antibody. An immunoassay for detecting antibodies directed against an 30 HCV antigen comprising incubating a sample suspected of containing anti-HCV antibodies with a probe polypeptide which contains an epitope of the HCV, under conditions which allow the formation of an antibody-antigen complex; and detecting the antibody-antigen complex containing the 35 probe antigen.

Also included in the invention are vaccines for treatment of HCV infection comprising an immunogenic peptide containing an HCV epitope, or an inactivated preparation of HCV, or an attenuated preparation of HCV.

5 Another aspect of the invention is a tissue culture grown cell infected with HCV.

Yet another aspect of the invention is a method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide 10 containing an HCV epitope in an amount sufficient to produce an immune response.

Still another aspect of the invention is a method for isolating cDNA derived from the genome of an unidentified infectious agent, comprising: (a) providing 15 host cells transformed with expression vectors containing a cDNA library prepared from nucleic acids isolated from tissue infected with the agent and growing said host cells under conditions which allow expression of polypeptide(s) encoded in the cDNA; (b) interacting the expression 20 products of the cDNA with an antibody containing body component of an individual infected with said infectious agent under conditions which allow an immunoreaction, and detecting antibody-antigen complexes formed as a result of the interacting; (c) growing host cells which express 25 polypeptides that form antibody-antigen complexes in step (b) under conditions which allow their growth as individual clones and isolating said clones; (d) growing cells from the clones of (c) under conditions which allow expression of polypeptide(s) encoded within the cDNA, and 30 interacting the expression products with antibody containing body components of individuals other than the individual in step (a) who are infected with the infectious agent and with control individuals uninfected with the agent, and detecting antibody-antigen complexes 35 formed as a result of the interacting; (e) growing host

cells which express polypeptides that form antibody-antigen complexes with antibody containing body components of infected individuals and individuals suspected of being infected, and not with said components of control 5 individuals, under conditions which allow their growth as individual clones and isolating said clones; and (f) isolating the cDNA from the host cell clones of (e).

Brief Description of the Drawings

10 Fig. 1 shows the double-stranded nucleotide sequence of the HCV cDNA insert in clone 5-1-1, and the putative amino acid sequence of the polypeptide encoded therein.

15 Fig. 2 shows the homologies of the overlapping HCV cDNA sequences in clones 5-1-1, 81, 1-2, and 91.

Fig. 3 shows a composite sequence of HCV cDNA derived from overlapping clones 81, 1-2, and 91, and the amino acid sequence encoded therein.

20 Fig. 4 shows the double-stranded nucleotide sequence of the HCV cDNA insert in clone 81, and the putative amino acid sequence of the polypeptide encoded therein.

25 Fig. 5 shows the HCV cDNA sequence in clone 36, the segment which overlaps the NANBV cDNA of clone 81, and the polypeptide sequence encoded within clone 36.

Fig. 6 shows the combined ORF of HCV cDNAs in clones 36 and 81, and the polypeptide encoded therein.

30 Fig. 7 shows the HCV cDNA sequence in clone 32, the segment which overlaps clone 81, and the polypeptide encoded therein.

Fig. 8 shows the HCV cDNA sequence in clone 35, the segment which overlaps clone 36, and the polypeptide encoded therein.

Fig. 9 shows the combined ORF of HCV cDNAs in clones 35, 36, 81, and 32, and the polypeptide encoded therein.

5 Fig. 10 shows the HCV cDNA sequence in clone 37b, the segment which overlaps clone 35, and the polypeptide encoded therein.

Fig. 11 shows the HCV cDNA sequence in clone 33b, the segment which overlaps clone 32, and the polypeptide encoded therein.

10 Fig. 12 shows the HCV cDNA sequence in clone 40b, the segment which overlaps clone 37b, and the polypeptide encoded therein.

Fig. 13 shows the HCV cDNA sequence in clone 25c, the segment which overlaps clone 33b, and the 15 polypeptide encoded therein.

Fig. 14 shows the nucleotide sequence and polypeptide encoded therein of the ORF which extends through the HCV cDNAs in clones 40b, 37b, 35, 36, 81, 32, 33b, and 25c.

20 Fig. 15 shows the HCV cDNA sequence in clone 33c, the segment which overlaps clones 40b and 33c, and the amino acids encoded therein.

Fig. 16 shows the HCV cDNA sequence in clone 8h, the segment which overlaps clone 33c, and the amino acids 25 encoded therein.

Fig. 17 shows the HCV cDNA sequence in clone 7e, the segment which overlaps clone 8h, and the amino acids encoded therein.

30 Fig. 18 shows the HCV cDNA sequence in clone 14c, the segment which overlaps clone 25c, and the amino acids encoded therein.

Fig. 19 shows the HCV cDNA sequence in clone 8f, the segment which overlaps clone 14c, and the amino acids encoded therein.

Fig. 20 shows the HCV cDNA sequence in clone 33f, the segment which overlaps clone 8f, and the amino acids encoded therein.

5 Fig. 21 shows the HCV cDNA sequence in clone 33g, the segment which overlaps clone 33f, and the amino acids encoded therein.

Fig. 22 shows the HCV cDNA sequence in clone 7f, the segment which overlaps the sequence in clone 7e, and the amino acids encoded therein.

10 Fig. 23 shows the HCV cDNA sequence in clone 11b, the segment which overlaps the sequence in clone 7f, and the amino acids encoded therein.

Fig. 24 shows the HCV cDNA sequence in clone 14i, the segment which overlaps the sequence in clone 11b, 15 and the amino acids encoded therein.

Fig. 25 shows the HCV cDNA sequence in clone 39c, the segment which overlaps the sequence in clone 33g, and the amino acids encoded therein.

Fig. 26 shows a composite HCV cDNA sequence 20 derived from the aligned cDNAs in clones 14i, 11b, 7f, 7e, 8h, 33c 40b 37b 35 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, and 39c also shown is the amino acid sequence of the polypeptide encoded in the extended ORF in the derived sequence.

25 Fig. 27 shows the sequence of the HCV cDNA in clone 12f, the segment which overlaps clone 14i, and the amino acids encoded therein.

Fig. 28 shows the sequence of the HCV cDNA in clone 35f, the segment which overlaps clone 39c, and the 30 amino acids encoded therein.

Fig. 29 shows the sequence of the HCV cDNA in clone 19g, the segment which overlaps clone 35f, and the amino acids encoded therein.

Fig. 30 shows the sequence of clone 26g, the segment which overlaps clone 19g, and the amino acids encoded therein.

5 Fig. 31 shows the sequence of clone 15e, the segment which overlaps clone 26g, and the amino acids encoded therein.

Fig. 32 shows the sequence in a composite cDNA, which was derived by aligning clones 12f through 15e in the 5' to 3' direction; it also shows the amino acids
10 encoded in the continuous ORF.

Fig. 33 shows a photograph of Western blots of a fusion protein, SOD-NANB₅₋₁₋₁, with chimpanzee serum from
15 chimpanzees infected with BB-NANB, HAV, and HBV.

Fig. 34 shows a photograph of Western blots of a fusion protein, SOD-NANB₅₋₁₋₁, with serum from humans infected with NANBV, HAV, HBV, and from control humans.

Fig. 35 is a map showing the significant
20 features of the vector pAB24.

Fig. 36 shows the putative amino acid sequence of the carboxy-terminus of the fusion polypeptide C100-3 and the nucleotide sequence encoding it.

Fig. 37A is a photograph of a coomassie blue
25 stained polyacrylamide gel which identifies C100-3 expressed in yeast.

Fig. 37B shows a Western blot of C100-3 with serum from a NANBV infected human.

Fig. 38 shows an autoradiograph of a Northern
30 blot of RNA isolated from the liver of a BB-NANBV infected chimpanzee, probed with BB-NANBV cDNA of clone 81.

Fig. 39 shows an autoradiograph of NANBV nucleic acid treated with RNase A or DNase I, and probed with BB-NANBV cDNA of clone 81.

Fig. 40 shows an autoradiograph of nucleic acids extracted from NANBV particles captured from infected plasma with anti-NANB₅₋₁₋₁, and probed with ³²P-labeled NANBV cDNA from clone 81.

5 Fig. 41 shows autoradiographs of filters containing isolated NANBV nucleic acids, probed with ³²P-labeled plus and minus strand DNA probes derived from NANBV cDNA in clone 81.

10 Fig. 42 shows the homologies between a polypeptide encoded in HCV cDNA and an NS protein from Dengue flavivirus.

Fig. 43 shows a histogram of the distribution of HCV infection in random samples, as determined by an ELISA screening.

15 Fig. 44 shows a histogram of the distribution of HCV infection in random samples using two configurations of immunoglobulin-enzyme conjugate in an ELISA assay.

Fig. 45 shows the sequences in a primer mix, derived from a conserved sequence in NS1 of flaviviruses.

Fig. 46 shows the HCV cDNA sequence in clone k9-1, the segment which overlaps the cDNA in Fig. 26, and the amino acids encoded therein.

Fig. 47 shows the sequence in a composite cDNA which was derived by aligning clones k9-1 through 15e in the 5' to 3' direction; it also shows the amino acids encoded in the continuous ORF.

Modes for Carrying Out the Invention

I. Definitions

The term "hepatitis C virus" has been reserved by workers in the field for an heretofore unknown etiologic agent of NANBH. Accordingly, as used herein, "hepatitis C virus" (HCV) refers to an agent causitive of NANBH, which was formerly referred to as NANBV and/or BB-NANBV. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein. As an extension of this terminology, the disease caused by HCV, formerly called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably herein.

The term "HCV", as used herein, denotes a viral species which causes NANBH, and attenuated strains or

defective interfering particles derived therefrom. As shown infra., the HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly on the 5 order of 10^{-3} to 10^{-4} per nucleotide (Fields & Knipe (1986)). Therefore, there are multiple strains within the HCV species described infra. The compositions and methods described herein, enable the propagation, identification, detection, and isolation of the various related strains.

10 Moreover, they also allow the preparation of diagnostics and vaccines for the various strains, and have utility in screening procedures for anti-viral agents for pharmacologic use in that they inhibit replication of HCV.

The information provided herein, although 15 derived from one strain of HCV, hereinafter referred to as CDC/HCV1, is sufficient to allow a viral taxonomist to identify other strains which fall within the species. As described herein, we have discovered that HCV is a Flavivirus or Flavi-like virus. The morphology and 20 composition of Flavivirus particles are known, and are discussed in Brinton (1986). Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

HCV encodes an epitope which is immunologically identifiable with an epitope in the HCV genome from which 30 the cDNAs described herein are derived; preferably the epitope is encoded in a cDNA described herein. The epitope is unique to HCV when compared to other known Flaviviruses. The uniqueness of the epitope may be determined by its immunological reactivity with HCV and 35 lack of immunological reactivity with other Flavivirus

species. Methods for determining immunological reactivity are known in the art, for example, by radioimmunoassay, by Elisa assay, by hemagglutination, and several examples of suitable techniques for assays are provided herein.

- 5 In addition to the above, the following parameters are applicable, either alone or in combination, in identifying a strain as HCV. Since HCV strains are evolutionarily related, it is expected that the overall homology of the genomes at the nucleotide level will be
- 10 about 40% or greater, preferably about 60% or greater, and even more preferably about 80% or greater; and in addition that there will be corresponding contiguous sequences of at least about 13 nucleotides. The correspondence between the putative HCV strain genomic sequence and the CDC/CH1
- 15 HCV cDNA sequence can be determined by techniques known in the art. For example, they can be determined by a direct comparison of the sequence information of the polynucleotide from the putative HCV, and the HCV cDNA sequence(s) described herein. For example, also, they can
- 20 be determined by hybridization of the polynucleotides under conditions which form stable duplexes between homologous regions (for example, those which would be used prior to S_1 digestion), followed by digestion with single stranded specific nuclease(s), followed by size
- 25 determination of the digested fragments.

Because of the evolutionary relationship of the strains of HCV, putative HCV strains are identifiable by their homology at the polypeptide level. Generally, HCV strains are more than about 40% homologous, preferably

30 more than about 60% homologous, and even more preferably more than about 80% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to

35 the sequences provided herein. For example also, the

nucleotide sequence of the genomic material of the putative HCV may be determined (usually via a cDNA intermediate); the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

5 As used herein, a polynucleotide "derived from" a designated sequence, for example, the HCV cDNA, particularly those exemplified in Figs. 1-32, or from an HCV genome, refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6
10 nucleotides, is preferably at least about 8 nucleotides, is more preferably at least about 10-12 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding, i.e., homologous to or complementary to, a region of the designated nucleotide sequence. Preferably,
15 the sequence of the region from which the polynucleotide is derived is homologous to or complementary to a sequence which is unique to an HCV genome. Whether or not a sequence is unique to the HCV genome can be determined by techniques known to those of skill in the art. For
20 example, the sequence can be compared to sequences in databanks, e.g., Genebank, to determine whether it is present in the uninfected host or other organisms. The sequence can also be compared to the known sequences of other viral agents, including those which are known to
25 induce hepatitis, e.g., HAV, HBV, and HDV, and to other members of the Flaviviridae. The correspondence or non-correspondence of the derived sequence to other sequences can also be determined by hybridization under the appropriate stringency conditions. Hybridization techniques
30 for determining the complementarity of nucleic acid sequences are known in the art, and are discussed infra. See also, for example, Maniatis et al. (1982). In addition, mismatches of duplex polynucleotides formed by hybridization can be determined by known techniques,
35 including for example, digestion with a nuclease such as

5 S1 that specifically digests single-stranded areas in duplex polynucleotides. Regions from which typical DNA sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

10 The derived polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription, which are based on the information provided by the sequence of bases in the region(s) from which the polynucleotide is derived. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be consistent with an intended use.

15 Similarly, a polypeptide or amino acid sequence "derived from" a designated nucleic acid sequence, for example, the sequences in Figs. 1-32, or from an HCV genome, refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence.

20 A recombinant or derived polypeptide is not necessarily translated from a designated nucleic acid sequence, for example, the sequences in Figs. 1-32, or from an HCV genome; it may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from mutated HCV.

25 The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its

origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature or in the form of a library; and/or (2) is linked to a polynucleotide other than that to which

5 it is linked in nature.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this
10 term includes double- and single-stranded DNA, as well as double- and single stranded RNA. It also includes modified, for example, by methylation and/or by capping, and unmodified forms of the polynucleotide.

As used herein, the term "HCV containing a
15 sequence corresponding to a cDNA" means that the HCV contains a polynucleotide sequence which is homologous to or complementary to a sequence in the designated DNA; the degree of homology or complementarity to the cDNA will be approximately 50% or greater, will preferably be at least
20 about 70%, and even more preferably will be at least about 90%. The sequences which correspond will be at least about 70 nucleotides, preferably at least about 80 nucleotides, and even more preferably at least about 90 nucleotides in length. The correspondence between the HCV
25 sequence and the cDNA can be determined by techniques known in the art, including, for example, a direct comparison of the sequenced material with the cDNAs described, or hybridization and digestion with single strand nucleases, followed by size determination of the
30 digested fragments.

The term "purified of viral polynucleotide"
refers to an HCV genome or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably
35 less than about 90% of polypeptides with which the viral

polynucleotide is naturally associated. Techniques for purifying viral polynucleotides from viral particles are known in the art, and include for example, disruption of the particle with a chaotropic agent, and separation of
5 the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

The term "purified viral polypeptided" refers to an HCV polypeptide or fragment thereof which is
10 essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90%, of cellular components with which the viral polypeptide is naturally associated. Techniques for purifying viral polypeptides are known in the art, and
15 examples of these techniques are discussed infra.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can
20 be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or
25 in genomic or total DNA complement as the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired
30 peptide, are included in the progeny intended by this definition, and are covered by the above terms.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, that behaves as an autonomous unit of polynucleotide replication within a
35 cell; i.e., capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers. The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not

limited to mRNA, cDNA, and recombinant polynucleotide sequences.

- "Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are
- 5 also present in and are unique to the designated polypeptide(s), usually HCV proteins. Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art, and are also ill-
- 10 lustrated infra. The uniqueness of an epitope can also be determined by computer searches of known data banks, e.g. Genebank, for the polynucleotide sequences which encode the epitope, and by amino acid sequence comparisons with other known proteins.
- 15 As used herein, "epitope" refers to an antigenic determinant of a polypeptide; an epitope could comprise 3 amino acids in a spatial conformation which is unique to the epitope, generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 20 8-10 such amino acids. Methods of determining the spatial conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance.

A polypeptide is "immunologically reactive" with

25 an antibody when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody binding, and/or by competition in binding using

30 as competitor(s) a known polypeptide(s) containing an epitope against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art.

As used herein, the term "immunogenic polypeptide containing an HCV epitope" includes naturally occurring HCV polypeptides or fragments thereof, as well as polypeptides prepared by other means, for example, 5 chemical synthesis, or the expression of the polypeptide in a recombinant organism.

The term "polypeptide" refers to a molecular chain of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and 10 proteins are included within the definition of polypeptide. This term also does not refer to post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. 15

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, or f-mating. The exogenous polynucleotide may be maintained as a non- 20 integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

"Treatment" as used herein refers to prophylaxis and/or therapy.

An "individual", as used herein, refers to 25 vertebrates, particularly members of the mammalian species, and includes but is not limited to domestic animals, sports animals, primates, and humans.

As used herein, the "plus strand" of a nucleic acid contains the sequence that encodes the polypeptide. 30 The "minus strand" contains a sequence which is complementary to that of the "plus strand".

As used herein, a "positive stranded genome" of a virus is one in which the genome, whether RNA or DNA, is single-stranded and which encodes a viral polypeptide(s). 35 Examples of positive stranded RNA viruses include

Togaviridae, Coronaviridae, Retroviridae, Picornaviridae, and Caliciviridae. Included also, are the Flaviviridae, which were formerly classified as Togaviradae. See Fields & Knipe (1986).

5 As used herein, "antibody containing body component" refers to a component of an individual's body which is a source of the antibodies of interest. Anti-body containing body components are known in the art, and include but are not limited to, for example, plasma, 10 serum, spinal fluid, lymph fluid, the external sections of the respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, white blood cells, and myelomas.

15 As used herein, "purified HCV" refers to a preparation of HCV which has been isolated from the cellular constituents with which the virus is normally associated, and from other types of viruses which may be present in the infected tissue. The techniques for isolating viruses are known to those of skill in the art, and include, for example, centrifugation and affinity 20 chromatography; a method of preparing purified HCV is discussed infra.

II. Description of the Invention

The practice of the present invention will 25 employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fisch & Sambrook, 30 MOLECULAR CLONING; A LABORATORY MANUAL (1982); DNA CLONING, VOLUMES I AND II (D.N. Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed. 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins 35 eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986);

IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and 5 M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN 10 PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986).

All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

The useful materials and processes of the present invention are made possible by the provision of a family of closely homologous nucleotide sequences isolated 20 from a cDNA library derived from nucleic acid sequences present in the plasma of an HCV infected chimpanzee. This family of nucleotide sequences is not of human or chimpanzee origin, since it hybridizes to neither human nor chimpanzee genomic DNA from uninfected individuals, 25 since nucleotides of this family of sequences are present only in liver and plasma of chimpanzees with HCV infection, and since the sequence is not present in Genebank. In addition, the family of sequences shows no significant homology to sequences contained within the HBV genome.

30 The sequence of one member of the family, contained within clone 5-1-1, has one continuous open reading frame (ORF) which encodes a polypeptide of approximately 50 amino acids. Sera from HCV infected humans contain antibodies which bind to this polypeptide, whereas 35 sera from non-infected humans do not contain antibodies to

this polypeptide. Finally, whereas the sera from uninfected chimpanzees do not contain antibodies to this polypeptide, the antibodies are induced in chimpanzees following acute NANBH infection. Moreover, antibodies to 5 this polypeptide are not detected in chimps and humans infected with HAV and HBV. By these criteria the sequence is a cDNA to a viral sequence, wherein the virus causes or is associated with NANBH; this cDNA sequence is shown in Fig. 1. As discussed infra, the cDNA sequence in clone 5-10 1-1 differs from that of the other isolated cDNAs in that it contains 28 extra base pairs.

A composite of other identified members of the cDNA family, which were isolated using as a probe a synthetic sequence equivalent to a fragment of the cDNA in 15 clone 5-1-1, is shown in Fig. 3. A member of the cDNA family which was isolated using a synthetic sequence derived from the cDNA in clone 81 is shown in Fig. 5, and the composite of this sequence with that of clone 81 is shown in Fig. 6. Other members of the cDNA family, 20 including those present in clones 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g and 15e are described in Section IV.A. A composite of the cDNAs in these clones is described in Section IV.A. 19, and shown in Fig. 32. The composite 25 cDNA shows that it contains one continuous ORF, and thus encodes a polyprotein.. This data is consistent with the suggestion, discussed infra., that HCV is a flavivirus or flavi-like virus.

The availability of this family of cDNAs shown 30 in Figs. 1-32, inclusive, permits the construction of DNA probes and polypeptides useful in diagnosing NANBH due to HCV infection and in screening blood donors as well as donated blood and blood products for infection. For example, from the sequences it is possible to synthesize

DNA oligomers of about 8-10 nucleotides, or larger, which are useful as hybridization probes to detect the presence of the viral genome in, for example, sera of subjects suspected of harboring the virus, or for screening donated blood for the presence of the virus. The family of cDNA sequences also allows the design and production of HCV specific polypeptides which are useful as diagnostic reagents for the presence of antibodies raised during NANBH. Antibodies to purified polypeptides derived from the cDNAs may also be used to detect viral antigens in infected individuals and in blood.

Knowledge of these cDNA sequences also enable the design and production of polypeptides which may be used as vaccines against HCV and also for the production of antibodies, which in turn may be used for protection against the disease, and/or for therapy of HCV infected individuals.

Moreover, the family of cDNA sequences enables further characterization of the HCV genome.

Polynucleotide probes derived from these sequences may be used to screen cDNA libraries for additional overlapping cDNA sequences, which, in turn, may be used to obtain more overlapping sequences. Unless the genome is segmented and the segments lack common sequences, this technique may be used to gain the sequence of the entire genome. However, if the genome is segmented, other segments of the genome can be obtained by repeating the lambda-gt11 serological screening procedure used to isolate the cDNA clones described herein, or alternatively by isolating the genome from purified HCV particles.

The family of cDNA sequences and the polypeptides derived from these sequences, as well as antibodies directed against these polypeptides are also useful in the isolation and identification of the BB-NANBV agent(s). For example, antibodies directed against HCV

epitopes contained in polypeptides derived from the cDNAs may be used in processes based upon affinity chromatography to isolate the virus. Alternatively, the antibodies may be used to identify viral particles

5 isolated by other techniques. The viral antigens and the genomic material within the isolated viral particles may then be further characterized.

The information obtained from further sequencing of the HCV genome(s), as well as from further

10 characterization of the HCV antigens and characterization of the genome enables the design and synthesis of additional probes and polypeptides and antibodies which may be used for diagnosis, for prevention, and for therapy of HCV induced NANBH, and for screening for infected blood

15 and blood-related products.

The availability of probes for HCV, including antigens and antibodies, and polynucleotides derived from the genome from which the family of cDNAs is derived also allows for the development of tissue culture systems which

20 will be of major use in elucidating the biology of HCV. This in turn, may lead to the development of new treatment regimens based upon antiviral compounds which preferentially inhibit the replication of, or infection by HCV.

25 The method used to identify and isolate the etiologic agent for NANBH is novel, and it may be applicable to the identification and/or isolation of heretofore uncharacterized agents which contain a genome, and which are associated with a variety of diseases, including

30 those induced by viruses, viroids, bacteria, fungi and parasites. In this method, a cDNA library was created from the nucleic acids present in infected tissue from an infected individual. The library was created in a vector which allowed the expression of polypeptides encoded in

35 the cDNA. Clones of host cells containing the vector,

which expressed an immunologically reactive fragment of a polypeptide of the etiologic agent, were selected by immunological screening of the expression products of the library with an antibody containing body component from

5 another individual previously infected with the putative agent. The steps in the immunological screening technique included interacting the expression products of the cDNA containing vectors with the antibody containing body component of a second infected individual, and detecting

10 the formation of antibody-antigen complexes between the expression product(s) and antibodies of the second infected individual. The isolated clones are screened further immunologically by interacting their expression products with the antibody containing body components of

15 other individuals infected with the putative agent and with control individuals uninfected with the putative agent, and detecting the formation of antigen-antibody complexes with antibodies from the infected individuals; and the cDNA containing vectors which encode polypeptides

20 which react immunologically with antibodies from infected individuals and individuals suspected of being infected with the agent, but not with control individuals are isolated. The infected individuals used for the construction of the cDNA library, and for the immunological

25 screening need not be of the same species.

The cDNAs isolated as a result of this method, and their expression products, and antibodies directed against the expression products, are useful in characterizing and/or capturing the etiologic agent. As described

30 in more detail infra, this method has been used successfully to isolate a family of cDNAs derived from the HCV genome.

II.A. Preparation of the cDNA Sequence

Pooled serum from a chimpanzee with chronic HCV infection and containing a high titer of the virus, i.e., at least 10^6 chimp infectious doses/ml (CID/ml) was used 5 to isolate viral particles; nucleic acids isolated from these particles was used as the template in the construction of a cDNA library to the viral genome. The procedures for isolation of putative HCV particles and for constructing the cDNA library in lambda-gt11 is discussed 10 in Section IV.A.1. Lambda-gt11 is a vector that has been developed specifically to express inserted cDNAs as fusion polypeptides with beta-galactosidase and to screen large numbers of recombinant phage with specific antisera raised against a defined antigen. The lambda-gt11 cDNA library 15 generated from a cDNA pool containing cDNA of approximate mean size of 200 base pairs was screened for encoded epitopes that could bind specifically with sera derived from patients who had previously experienced NANB hepatitis. Huynh, T.V. et al. (1985). Approximately 10^6 20 phages were screened, and five positive phages were identified, purified, and then tested for specificity of binding to sera from different humans and chimpanzees previously infected with the HCV agent. One of the phages, 5-1-1, bound 5 of the 8 human sera tested. This 25 binding appeared selective for sera derived from patients with prior NANB hepatitis infections since 7 normal blood donor sera did not exhibit such binding.

The sequence of the cDNA in recombinant phage 5-1-1 was determined, and is shown in Fig. 1. The 30 polypeptide encoded by this cloned cDNA, which is in the same translational frame as the N-terminal beta-Galactosidase moiety of the fusion polypeptide is shown above the nucleotide sequence. This translational ORF, therefore, encodes an epitope(s) specifically recognized 35 by sera from patients with NANB hepatitis infections.

The availability of the cDNA in recombinant phage 5-1-1 has allowed for the isolation of other clones containing additional segments and/or alternative segments of cDNA to the viral genome. The lambda-gt11 cDNA library 5 described supra, was screened using a synthetic polynucleotide derived from the sequence of the cloned 5-1-1 cDNA. This screening yielded three other clones, which were identified as 81, 1-2 and 91; the cDNAs contained within these clones were sequenced. See Sections 10 IV.A.3. and IV.A.4. The homologies between the four independent clones are shown in Fig. 2, where the homologies are indicated by the vertical lines. Sequences of nucleotides present uniquely in clones 5-1-1, 81, and 91 are indicated by small letters.

15 The cloned cDNAs present in recombinant phages in clones 5-1-1, 81, 1-2, and 91 are highly homologous, and differ in only two regions. First, nucleotide number 67 in clone 1-2 is a thymidine, whereas the other three clones contain a cytidine residue in this position. This 20 substitution, however, does not alter the nature of the encoded amino acid.

The second difference between the clones is that clone 5-1-1 contains 28 base pairs at its 5'-terminus which are not present in the other clones. The extra 25 sequence may be a 5'-terminal cloning artifact; 5'- terminal cloning artifacts are commonly observed in the products of cDNA methods.

Synthetic sequences derived from the 5'-region and the 3'-region of the HCV cDNA in clone 81 were used to 30 screen and isolate cDNAs from the lambda-gt11 NANBV cDNA library, which overlapped clone 81 cDNA (Section IV.A.5.). The sequences of the resulting cDNAs, which are in clone 36 and clone 32, respectively, are shown in Fig. 5 and Fig. 7.

Similarly, a synthetic polynucleotide based on the 5'-region of clone 36 was used to screen and isolate cDNAs from the lambda gt-11 NANBV cDNA library which overlapped clone 36 cDNA (Section IV.A.8.). A purified 5 clone of recombinant phage-containing cDNA which hybridized to the synthetic polynucleotide probe was named clone 35 and the NANBV cDNA sequence contained within this clone is shown in Fig. 8.

By utilizing the technique of isolating overlapping 10 cDNA sequences, clones containing additional upstream and downstream HCV cDNA sequences have been obtained. The isolation of these clones, is described infra in Section IV.A.

Analysis of the nucleotide sequences of the HCV 15 cDNAs encoded within the isolated clones show that the composite cDNA contains one long continuous ORF. Fig. 26 shows the sequence of the composite cDNA from these clones, along with the putative HCV polypeptide encoded therein.

20 The description of the method to retrieve the cDNA sequences is mostly of historical interest. The resultant sequences (and their complements) are provided herein, and the sequences, or any portion thereof, could be prepared using synthetic methods, or by a combination 25 of synthetic methods with retrieval of partial sequences using methods similar to those described herein.

Lambda-gt11 strains replicated from the HCV cDNA library and from clones 5-1-1, 81, 1-2 and 91 have been deposited under the terms of the Budapest Treaty with the 30 American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

<u>lambda-gt11</u>	<u>ATCC No.</u>	<u>Deposit Date</u>
HCV cDNA library	40394	1 Dec. 1987
clone 81	40388	17 Nov. 1987
clone 91	40389	17 Nov. 1987
5 clone 1-2	40390	17 Nov. 1987
clone 5-1-1	40391	18 Nov. 1987

The designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. These deposits and other deposited materials mentioned herein are intended for convenience only, and are not required to practice the present invention in view of the description here. The HCV cDNA sequences in all of the deposited materials are incorporated herein by reference.

The description above, of "walking" the genome by isolating overlapping cDNA sequences from the HCV lambda gt-11 library provides one method by which cDNAs corresponding to the entire HCV genome may be isolated. However, given the information provided herein, other methods for isolating these cDNAs are obvious to one of skill in the art. Some of these methods are described in Section IV.A., infra.

25 II.B. Preparation of Viral Polypeptides and Fragments

The availability of cDNA sequences, either those isolated by utilizing the cDNA sequences in Figs. 1-32, as discussed infra, as well as the cDNA sequences in these

figures, permits the construction of expression vectors encoding antigenically active regions of the polypeptide encoded in either strand. These antigenically active regions may be derived from coat or envelope antigens or 5 from core antigens, including, for example, polynucleotide binding proteins, polynucleotide polymerase(s), and other viral proteins required for the replication and/or assembly of the virus particle. Fragments encoding the desired polypeptides are derived from the cDNA clones 10 using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-Galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the 15 production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986. Vectors encoding fusion polypeptides of SOD and HCV polypeptides, i.e., NANB₅₋₁₋₁, NANB₈₁, and C100-3, which is encoded in a 20 composite of HCV cDNAs, are described in Sections IV.B.1, IV.B.2, and IV.B.4, respectively. Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be obtained as a recombinant polypeptide, such as a mature or fusion protein; 25 alternatively, a polypeptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be 30 ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell lines is given in Section 35 III.A., infra. The polypeptide is then isolated from

lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such polypeptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, as discussed in Section II.J. herein below, antibodies to these polypeptides are useful for isolating and identifying HCV particles.

The HCV antigens may also be isolated from HCV virions. The virions may be grown in HCV infected cells in tissue culture, or in an infected host.

II.C. Preparation of Antigenic Polypeptides and Conjugation with Carrier

An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of HCV antigen. Accordingly, using the cDNAs of HCV as a basis, DNAs encoding short segments of HCV polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the synthesized polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridyl-thio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun. Rev. (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles, see, for example, section II.D. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin,

tetanus toxoid, and other proteins well known to those skilled in the art.

II.D. Preparation of Hybrid Particle Immunogens Containing HCV Epitopes

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. Constructs wherein the NANBV epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S. cerevisiae (Valenzuela et al. (1982)), as well as in, for example, mammalian cells (Valenzuela, P., et al. (1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1986. Both applications are assigned to the herein assignee, and are incorporated herein by reference. These constructs may also be expressed in mammalian cells such as Chinese

hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

10

II.E. Preparation of Vaccines

Vaccines may be prepared from one or more immunogenic polypeptides derived from HCV cDNA as well as from the cDNA sequences in the Figs. 1-32, or from the HCV genome to which they correspond. The observed homology between HCV and Flaviviruses provides information concerning the polypeptides which are likely to be most effective as vaccines, as well as the regions of the genome in which they are encoded. The general structure of the Flavivirus genome is discussed in Rice et al (1986). The flavivirus genomic RNA is believed to be the only virus-specific mRNA species, and it is translated into the three viral structural proteins, i.e., C, M, and E, as well as two large nonstructural proteins, NV4 and NV5, and a complex set of smaller nonstructural proteins. It is known that major neutralizing epitopes for Flaviviruses reside in the E (envelope) protein (Roehrig (1986)). The corresponding HCV E gene and polypeptide encoding region can be predicted, based upon the homology to Flaviviruses. Thus, vaccines may be comprised of recombinant polypeptides containing epitopes of HCV E. These polypeptides may be expressed in bacteria, yeast, or mammalian cells, or alternatively may be isolated from viral preparations. It is also anticipated that the other structural proteins may also contain epitopes which give

rise to protective anti-HCV antibodies. Thus, polypeptides containing the epitopes of E, C, and M may also be used, whether singly or in combination, in HCV vaccines.

- 5 In addition to the above, it has been shown that immunization with NS1 (nonstructural protein 1), results in protection against yellow fever (Schlesinger et al (1986)). This is true even though the immunization does not give rise to neutralizing antibodies. Thus,
- 10 particularly since this protein appears to be highly conserved among Flaviviruses, it is likely that HCV NS1 will also be protective against HCV infection. Moreover, it also shows that nonstructural proteins may provide protection against viral pathogenicity, even if they do
- 15 not cause the production of neutralizing antibodies.

In view of the above, multivalent vaccines against HCV may be comprised of one or more structural proteins, and/or one or more nonstructural proteins.

- These vaccines may be comprised of, for example,
- 20 recombinant HCV polypeptides and/or polypeptides isolated from the virions. In addition, it may be possible to use inactivated HCV in vaccines; inactivation may be by the preparation of viral lysates, or by other means known in the art to cause inactivation of Flaviviruses, for
- 25 example, treatment with organic solvents or detergents, or treatment with formalin. Moreover, vaccines may also be prepared from attenuated HCV strains. The preparation of attenuated HCV strains is described infra.

- It is known that some of the proteins in
- 30 Flaviviruses contain highly conserved regions, thus, some immunological cross-reactivity is expected between HCV and other Flaviviruses. It is possible that shared epitopes between the Flaviviruses and HCV will give rise to protective antibodies against one or more of the disorders
- 35 caused by these pathogenic agents. Thus, it may be

possible to design multipurpose vaccines based upon this knowledge.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include
5 suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of
10 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions
15 take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with
20 the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.
25

30

II.F. Dosage and Administration of Vaccines

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally
35

in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of 5 active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A 10 multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a 15 subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic HCV antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, immune globulins. 20

II.G. Preparation of Antibodies Against HCV Epitopes

25 The immunogenic polypeptides prepared as described above are used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide 30 bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity 35 chromatography. Techniques for producing and processing

polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Alternatively, polyclonal antibodies may be isolated from a mammal which has been previously infected 5 with HCV. An example of a method for purifying antibodies to HCV epitopes from serum from an infected individual, based upon affinity chromatography and utilizing a fusion polypeptide of SOD and a polypeptide encoded within cDNA clone 5-1-1, is presented in Section V.E.

10 Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or 15 transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 20 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, 25 which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

30 Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985).

Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for 5 treatment of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

II.H. Diagnostic Oligonucleotide Probes and Kits

Using the disclosed portions of the isolated HCV 10 cDNAs as a basis, including those in Figs. 1-32, oligomers of approximately 8 nucleotides or more can be prepared, either by excision or synthetically, which hybridize with the HCV genome and are useful in identification of the viral agent(s), further characterization of the viral 15 genome(s), as well as in detection of the virus(es) in diseased individuals. The probes for HCV polynucleotides (natural or derived) are a length which allows the detection of unique viral sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences 20 of 10-12 nucleotides are preferred, and about 20 nucleotides appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. Among useful 25 probes, for example, are the clone 5-1-1 and the additional clones disclosed herein, as well as the various oligomers useful in probing cDNA libraries, set forth below. A complement to any unique portion of the HCV genome will be satisfactory. For use as probes, complete 30 complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

For use of such probes as diagnostics, the biological sample to be analyzed, such as blood or serum, is treated, if desired, to extract the nucleic acids 35 contained therein. The resulting nucleic acid from the

sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. The probes are then labeled. Suitable labels, and methods 5 for labeling probes are known in the art, and include, for example, radioactive labels incorporated by nick translation or kinasing, biotin, fluorescent probes, and chemiluminescent probes. The nucleic acids extracted from the sample are then treated with the labeled probe under 10 hybridization conditions of suitable stringencies.

The probes can be made completely complementary to the HCV genome. Therefore, usually high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency should 15 only be used if the probes are complementary to regions of the viral genome which lack heterogeneity. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and 20 concentration of formamide. These factors are outlined in, for example, Maniatis, T. (1982).

Generally, it is expected that the HCV genome sequences will be present in serum of infected individuals at relatively low levels, i.e., at approximately 10^2 - 10^3 25 sequences per ml. This level may require that amplification techniques be used in hybridization assays. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT- 30 tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT application 84/03520 and EPA124221 describe a DNA hybridization assay in which: (1) analyte is annealed to a single-stranded DNA probe that is 35 complementary to an enzyme-labeled oligonucleotide; and

(2) the resulting tailed duplex is hybridized to an enzyme-labeled oligonucleotide. EPA 204510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an 5 amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands. A particularly desirable technique may first involve amplification of the target HCV sequences in sera 10 approximately 10,000 fold, i.e., to approximately 10^6 sequences/ml. This may be accomplished, for example, by the technique of Saiki et al. (1986). The amplified sequence(s) may then be detected using a hybridization assay which is described in co-pending U.S. Application, 15 Attorney Docket No. 2300-0171, which was filed 15 October 1987, is assigned to the herein assignee, and is hereby incorporated herein by reference. This hybridization assay, which should detect sequences at the level of 10^6 /ml utilizes nucleic acid multimers which bind to single- 20 stranded analyte nucleic acid, and which also bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO 225,807, 25 published June 16, 1987, which is assigned to the herein assignee, and which is hereby incorporated herein by reference.

The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be 30 labeled; alternatively, the probe DNA may be unlabeled and the ingredients for labeling may be included in the kit. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, as well as instructions 35 for conducting the test.

II.I. Immunoassay and Diagnostic Kits

Both the polypeptides which react immuno-logically with serum containing HCV antibodies, for example, those derived from or encoded within the clones 5 described in Section IV.A., and composites thereof, (see section IV.A.) and the antibodies raised against the HCV specific epitopes in these polypeptides, see for example Section IV.E, are useful in immunoassays to detect presence of HCV antibodies, or the presence of the virus 10 and/or viral antigens, in biological samples, including for example, blood or serum samples. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. For example, the immunoassay may utilize one viral antigen, for example, a 15 polypeptide derived from any of the clones containing HCV cDNA described in Section IV.A., or from the composite cDNAs derived from the cDNAs in these clones, or from the HCV genome from which the cDNA in these clones is derived; alternatively, the immunoassay may use a combination of 20 viral antigens derived from these sources. It may use, for example, a monoclonal antibody directed towards a viral epitope(s), a combination of monoclonal antibodies directed towards one viral antigen, monoclonal antibodies directed towards different viral antigens, polyclonal 25 antibodies directed towards the same viral antigen, or polyclonal antibodies directed towards different viral antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or 30 may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays 35

which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The Flavivirus model for HCV allows predictions regarding the likely location of diagnostic epitopes for the virion structural proteins. The C, pre-M, M, and E domains are all likely to contain epitopes of significant potential for detecting viral antigens, and particularly for diagnosis. Similarly, domains of the nonstructural proteins are expected to contain important diagnostic epitopes (e.g., NS5 encoding a putative polymerase; and NS1 encoding a putative complement-binding antigen). Recombinant polypeptides, or viral polypeptides, which include epitopes from these specific domains may be useful for the detection of viral antibodies in infectious blood donors and infected patients.

In addition, antibodies directed against the E and/or M proteins can be used in immunoassays for the detection of viral antigens in patients with HCV caused NANBH, and in infectious blood donors. Moreover, these antibodies will be extremely useful in detecting acute-phase donors and patients.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing HCV epitopes or antibodies directed against HCV epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay, as well as a suitable set of assay instructions.

II.J. Further Characterization of the HCV Genome, Virions, and Viral Antigens Using Probes Derived From cDNA to the Viral Genome

The HCV cDNA sequence information in the clones described in Section IV.A., as shown in Figs. 1-32,

inclusive, may be used to gain further information on the sequence of the HCV genome, and for identification and isolation of the HCV agent, and thus will aid in its characterization including the nature of the genome, the 5 structure of the viral particle, and the nature of the antigens of which it is composed. This information, in turn, can lead to additional polynucleotide probes, polypeptides derived from the HCV genome, and antibodies directed against HCV epitopes which would be useful for 10 the diagnosis and/or treatment of HCV caused NANBH.

The cDNA sequence information in the above-mentioned clones is useful for the design of probes for the isolation of additional cDNA sequences which are derived from as yet undefined regions of the HCV genome(s) 15 from which the cDNAs in clones described in Section IV.A. are derived. For example, labeled probes containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the 5'-termini or 3'-termini of the family of HCV cDNA sequences shown in Figs. 1, 3, 6, 9, 14 20 and 32 may be used to isolate overlapping cDNA sequences from HCV cDNA libraries. These sequences which overlap the cDNAs in the above-mentioned clones, but which also contain sequences derived from regions of the genome from 25 which the cDNA in the above mentioned clones are not derived, may then be used to synthesize probes for identification of other overlapping fragments which do not necessarily overlap the cDNAs in the clones described in Section IV.A. Unless the HCV genome is segmented and the 30 segments lack common sequences, it is possible to sequence the entire viral genome(s) utilizing the technique of isolation of overlapping cDNAs derived from the viral genome(s). Although it is unlikely, if the genome is a 35 segmented genome which lacks common sequences, the sequence of the genome can be determined by serologically

screening lambda-gt11 HCV cDNA libraries, as used to isolate clone 5-1-1, sequencing cDNA isolates, and using the isolated cDNAs to isolate overlapping fragments, using the technique described for the isolation and sequencing 5 of the clones described in Section IV.A. Alternatively, characterization of the genomic segments could be from the viral genome(s) isolated from purified HCV particles. Methods for purifying HCV particles and for detecting them during the purification procedure are described herein, 10 infra. Procedures for isolating polynucleotide genomes from viral particles are known in the art, and one procedure which may be used is shown in Example IV.A.1. The isolated genomic segments could then be cloned and sequenced. Thus, with the information provided herein, it 15 is possible to clone and sequence the HCV genome(s) irrespective of their nature.

Methods for constructing cDNA libraries are known in the art, and are discussed supra and infra; a method for the construction of HCV cDNA libraries in 20 lambda-gt11 is discussed infra in Section IV.A. However, cDNA libraries which are useful for screening with nucleic acid probes may also be constructed in other vectors known in the art, for example, lambda-gt10 (Huynh et al. (1985)). The HCV derived cDNA detected by the probes 25 derived from the cDNAs in Figs. 1-32, and from the probes synthesized from polynucleotides derived from these cDNAs, may be isolated from the clone by digestion of the isolated polynucleotide with the appropriate restriction enzyme(s), and sequenced. See, for example, Section 30 IV.A.3. and IV.A.4. for the techniques used for the isolation and sequencing of HCV cDNA which overlaps HCV cDNA in clone 5-1-1, Sections IV.A.5-IV.A.7 for the isolation and sequencing of HCV cDNA which overlaps that in clone 81, and Section IV.A.8 and IV.A.9 for the

isolation and sequencing of a clone which overlaps another clone (clone 36), which overlaps clone 81.

The sequence information derived from these overlapping HCV cDNAs is useful for determining areas of homology and heterogeneity within the viral genome(s), which could indicate the presence of different strains of the genome, and/or of populations of defective particles. It is also useful for the design of hybridization probes to detect HCV or HCV antigens or HCV nucleic acids in biological samples, and during the isolation of HCV (discussed infra), utilizing the techniques described in Section II.G. Moreover, the overlapping cDNAs may be used to create expression vectors for polypeptides derived from the HCV genome(s) which also encode the polypeptides encoded in clones 5-1-1, 36, 81, 91, and 1-2, and in the other clones described in Section IV.A. The techniques for the creation of these polypeptides containing HCV epitopes, and for antibodies directed against HCV epitopes contained within them, as well as their uses, are analogous to those described for polypeptides derived from NANBV cDNA sequences contained within clones 5-1-1, 32, 35, 36, 1-2, 81, and 91, discussed supra and infra.

Encoded within the family of cDNA sequences contained within clones 5-1-1, 32, 35, 36, 81, 91, 1-2, and the other clones described in Section IV.A. are antigen(s) containing epitopes which appear to be unique to HCV; i.e., antibodies directed against these antigens are absent from individuals infected with HAV or HBV, and from individuals not infected with HCV (see the serological data presented in Section IV.B.). Moreover, a comparison of the sequence information of these cDNAs with the sequences of HAV, HBV, HDV, and with the genomic sequences in Genebank indicates that minimal homology exists between these cDNAs and the polynucleotide sequences of those sources. Thus, antibodies directed against the

antigens encoded within the cDNAs of these clones may be used to identify BB-NANBV particles isolated from infected individuals. In addition, they are also useful for the isolation of NANBH agent(s).

- 5 HCV particles may be isolated from the sera from BB-NANBV infected individuals or from cell cultures by any of the methods known in the art, including for example, techniques based on size discrimination such as sedimentation or exclusion methods, or techniques based on density
- 10 such as ultracentrifugation in density gradients, or precipitation with agents such as polyethylene glycol, or chromatography on a variety of materials such as anionic or cationic exchange materials, and materials which bind due to hydrophobicity, as well as affinity columns. During
- 15 the isolation procedure the presence of HCV may be detected by hybridization analysis of the extracted genome, using probes derived from the HCV cDNAs described supra, or by immunoassay (see Section II.I.) utilizing as probes antibodies directed against HCV antigens encoded
- 20 within the family of cDNA sequences shown in Figs. 1-32, and also directed against HCV antigens encoded within the overlapping HCV cDNA sequences discussed supra. The antibodies may be monoclonal, or polyclonal, and it may be desirable to purify the antibodies before their use in the
- 25 immunoassay. A purification procedure for polyclonal antibodies directed against antigen(s) encoded within clone 5-1-1 is described in Section IV.E; analogous purification procedures may be utilized for antibodies directed against other HCV antigens.
- 30 Antibodies directed against HCV antigens encoded within the family of cDNAs shown in Figs. 1-32, as well as those encoded within overlapping HCV cDNAs, which are affixed to solid supports are useful for the isolation of HCV by immunoaffinity chromatography. Techniques for
- 35 immunoaffinity chromatography are known in the art,

including techniques for affixing antibodies to solid supports so that they retain their immunoselective activity; the techniques may be those in which the antibodies are adsorbed to the support (see, for example, Kurstak in ENZYME IMMUNODIAGNOSIS, page 31-37), as well as those in which the antibodies are covalently linked to the support. Generally, the techniques are similar to those used for covalent linking of antigens to a solid support, which are generally described in Section II.C.; however, spacer groups may be included in the bifunctional coupling agents so that the antigen binding site of the antibody remains accessible.

During the purification procedure the presence of HCV may be detected and/or verified by nucleic acid hybridization, utilizing as probes polynucleotides derived from the family of HCV cDNA sequences shown in Figs. 1-32, as well as from overlapping HCV cDNA sequences, described supra. In this case, the fractions are treated under conditions which would cause the disruption of viral particles, for example, with detergents in the presence of chelating agents, and the presence of viral nucleic acid determined by hybridization techniques described in Section II.H. Further confirmation that the isolated particles are the agents which induce HCV may be obtained by infecting chimpanzees with the isolated virus particles, followed by a determination of whether the symptoms of NANBH result from the infection.

Viral particles from the purified preparations may then be further characterized. The genomic nucleic acid has been purified. Based upon its sensitivity to RNase, and not DNase I, it appears that the virus is composed of an RNA genome. See Example IV.C.2., infra. The strandedness and circularity or non-circularity can be determined by techniques known in the art, including, for example, its visualization by electron microscopy, its

migration in density gradients, and its sedimentation characteristics. Based upon the hybridization of the captured HCV genome to the negative strands of HCV cDNAs, it appears that HCV may be comprised of a positive
5 stranded RNA genome (see Section IV.H.1). Techniques such as these are described in, for example, METHODS IN ENZYMOLOGY. In addition, the purified nucleic acid can be cloned and sequenced by known techniques, including reverse transcription since the genomic material is RNA.
10 See, for example, Maniatis (1982), and Glover (1985). Utilizing the nucleic acid derived from the viral particles, it is possible to sequence the entire genome, whether or not it is segmented.

Examination of the homology of the polypeptide
15 encoded within the continuous ORF of combined clones 14i through 39c (see Fig. 26), shows that the HCV polypeptide contains regions of homology with the corresponding proteins in conserved regions of flaviviruses. An example of this is described in Section IV.H.3. This finding has
20 many important ramifications. First, this evidence, in conjunction with the results which show that HCV contains a positive-stranded genome, the size of which is approximately 10,000 nucleotides, is consistent with the suggestion that HCV is a flavivirus, or flavi-like virus.
25 Generally, flavivirus virions and their genomes have a relatively consistent structure and organization, which are known. See Rice et al. (1986), and Brinton, M.A. (1988). Thus, the structural genes encoding the polypeptides C, pre-M/M, and E may be located in the 5'-
30 terminus of the genome upstream of clone 14i. Moreover, using the comparison with other flaviviruses, predictions as to the precise location of the sequences encoding these proteins can be made.

Isolation of the sequences upstream of those in
35 clone 14i may be accomplished in a number of ways which,

given the information herein, would be obvious to one of skill in the art. For example, the genome "walking" technique, may be used to isolate other sequences which are 5' to those in clone 14i, but which overlap that 5 clone; this in turn leads to the isolation of additional sequences. This technique has been amply demonstrated infra, in Section IV.A.. For example, also, it is known that the flaviviruses have conserved epitopes and regions of conserved nucleic acid sequences. Polynucleotides 10 containing the conserved sequences may be used as probes which bind the HCV genome, thus allowing its isolation. In addition, these conserved sequences, in conjunction with those derived from the HCV cDNAs shown in Fig. 22, may be used to design primers for use in systems which 15 amplify the genome sequences upstream of those in clone 14i, using polymerase chain reaction technology. An example of this is described infra.

The structure of the HCV may also be determined and its components isolated. The morphology and size may 20 be determined by, for example, electron microscopy. The identification and localization of specific viral polypeptide antigens such as coat or envelope antigens, or internal antigens, such as nucleic acid binding proteins, core antigens, and polynucleotide polymerase(s) may also 25 be determined by, for example, determining whether the antigens are present as major or minor viral components, as well as by utilizing antibodies directed against the specific antigens encoded within isolated cDNAs as probes. This information is useful in the design of vaccines; for 30 example, it may be preferable to include an exterior antigen in a vaccine preparation. Multivalent vaccines may be comprised of, for example, a polypeptide derived from the genome encoding a structural protein, for example, E, as well as a polypeptide from another portion

of the genome, for example, a nonstructural or structural polypeptide.

II.K. Cell Culture Systems and Animal Model Systems for

5 HCV Replication

The suggestion that HCV is a flavivirus or flavi-like virus also provides information on methods for growing HCV. The term "flavi-like" means that the virus shows a significant amount of homology to the known 10 conserved regions of flaviviruses and that the majority of the genome is a single ORF. Methods for culturing flaviviruses are known to those of skill in the art (See, for example, the reviews by Brinton (1986) and Stollar, V. (1980)). Generally, suitable cells or cell lines for 15 culturing HCV may include those known to support Flavivirus replication, for example, the following: monkey kidney cell lines (e.g. MK₂, VERO); porcine kidney cell lines (e.g. PS); baby hamster kidney cell lines (e.g. BHK); murine macrophage cell lines (e.g., P388D1, MK1, 20 Mml); human macrophage cell lines (e.g., U-937); human peripheral blood leukocytes; human adherent monocytes; hepatocytes or hepatocyte cell lines (e.g., HUH7, HEPG2); embryos or embryonic cells (e.g., chick embryo fibroblasts); or cell lines derived from invertebrates, 25 preferably from insects (e.g. drosophila cell lines), or more preferably from arthropods, for example, mosquito cell lines (e.g., A. Albopictus, Aedes aegypti, Cutex tritaeniorhynchus) or tick cell lines (e.g. RML-14 Dermacentor parumapertus). 30 It is possible that primary hepatocytes can be cultured, and then infected with HCV; or alternatively, the hepatocyte cultures could be derived from the livers of infected individuals (e.g., humans or chimpanzees). The latter case is an example of a cell which is infected 35 in vivo being passed in vitro. In addition, various

immortalization methods can be used to obtain cell-lines derived from hepatocyte cultures. For example, primary liver cultures (before and after enrichment of the hepatocyte population) may be fused to a variety of cells to maintain stability. For example, also, cultures may be infected with transforming viruses, or transfected with transforming genes in order to create permanent or semipermanent cell lines. In addition, for example, cells in liver cultures may be fused to established cell lines (e.g., HepG2). Methods for cell fusion are known in the art, and include, for example, the use of fusion agents such as polyethylene glycol, Sendai Virus, and Epstein-Barr virus.

As discussed above, HCV is a Flavivirus or Flavi-like virus. Therefore, it is probable that HCV infection of cell lines may be accomplished by techniques known in the art for infecting cells with Flaviviruses. These include, for example, incubating the cells with viral preparations under conditions which allow viral entry into the cell. In addition, it may be possible to obtain viral production by transfecting the cells with isolated viral polynucleotides. It is known that Togavirus and Flavivirus RNAs are infectious in a variety of vertebrate cell lines (Pfefferkorn and Shapiro (1974)), and in a mosquito cell line (Peleg (1969)).

Methods for transfecting tissue culture cells with RNA duplexes, positive stranded RNAs, and DNAs (including cDNAs) are known in the art, and include, for example, techniques which use electroporation, and precipitation with DEAE-Dextran or calcium phosphate. An abundant source of HCV RNA can be obtained by performing in vitro transcription of an HCV cDNA corresponding to the complete genome. Transfection with this material, or with cloned HCV cDNA should result in viral replication and the in vitro propagation of the virus.

In addition to cultured cells, animal model systems may be used for viral replication; animal systems in which flaviviruses are known to those of skill in the art (See, for example, the review by Monath (1986)).
Thus, HCV replication may occur not only in chimpanzees,
5 but also in, for example, marmosets and suckling mice.

III.L. Screening for Anti-Viral Agents for HCV

The availability of cell culture and animal model systems for HCV also makes possible screening for
10 anti-viral agents which inhibit HCV replication, and particularly for those agents which preferentially allow cell growth and multiplication while inhibiting viral replication. These screening methods are known by those of skill in the art. Generally, the anti-viral agents are
15 tested at a variety of concentrations, for their effect on preventing viral replication in cell culture systems which support viral replication, and then for an inhibition of infectivity or of viral pathogenicity (and a low level of toxicity) in an animal model system.
20

The methods and compositions provided herein for detecting HCV antigens and HCV polynucleotides are useful for screening of anti-viral agents in that they provide an alternative, and perhaps more sensitive means, for detecting the agent's effect on viral replication than the
25 cell plaque assay or ID₅₀ assay. For example, the HCV-polynucleotide probes described herein may be used to quantitate the amount of viral nucleic acid produced in a cell culture. This could be accomplished, for example, by hybridization or competition hybridization of the infected
30 cell nucleic acids with a labeled HCV-polynucleotide probe. For example, also, anti-HCV antibodies may be used to identify and quantitate HCV antigen(s) in the cell culture utilizing the immunoassays described herein. In addition, since it may be desirable to quantitate HCV
35

antigens in the infected cell culture by a competition assay, the polypeptides encoded within the HCV cDNAs described herein are useful in these competition assays. Generally, a recombinant HCV polypeptide derived from the 5 HCV cDNA would be labeled, and the inhibition of binding of this labeled polypeptide to an HCV polypeptide due to the antigen produced in the cell culture system would be monitored. Moreover, these techniques are particularly useful in cases where the HCV may be able to replicate in 10 a cell line without causing cell death.

II.M. Preparation of Attenuated Strains of HCV

In addition to the above, utilizing the tissue culture systems and/or animal model systems, it may be 15 possible to isolate attenuated strains of HCV. These strains would be suitable for vaccines, or for the isolation of viral antigens. Attenuated strains are isolatable after multiple passages in cell culture and/or an animal model. Detection of an attenuated strain in an infected 20 cell or individual is achievable by techniques known in the art, and could include, for example, the use of antibodies to one or more epitopes encoded in HCV as a probe or the use of a polynucleotide containing an HCV sequence of at least about 8 nucleotides as a probe. 25 Alternatively, or in addition, an attenuated strain may be constructed utilizing the genomic information of HCV provided herein, and utilizing recombinant techniques. Generally, one would attempt to delete a region of the genome encoding, for example, a polypeptide related to 30 pathogenicity, but which allows viral replication. In addition, the genome construction would allow the expression of an epitope which gives rise to neutralizing antibodies for HCV. The altered genome could then be utilized to transform cells which allow HCV replication, 35 and the cells grown under conditions to allow viral

replication. Attenuated HCV strains are useful not only for vaccine purposes, but also as sources for the commercial production of viral antigens, since the processing of these viruses would require less stringent protection measures for the employees involved in viral 5 production and/or the production of viral products.

III. General Methods

The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, 10 sequencing clones, constructing expression vectors, transforming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. 15 However, as a general guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

III.A. Hosts and Expression Control Sequences

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, E. coli is most frequently used. Expression control 20 sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid 25 containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful 30 transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase 35

(penicillinase) and lactose promoter systems (Chang et al. (1977)), the tryptophan (trp) promoter system (Goeddel et al. (1980)) and the lambda-derived P_L promoter and N gene ribosome binding site (Shimatake et al. (1981)) and the hybrid tac promoter (De Boer et al. (1983)) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with E. coli; if desired, other prokaryotic hosts such as strains of Bacillus or Pseudomonas may be used, with corresponding control sequences.

Eukaryotic hosts include yeast and mammalian cells in culture systems. Saccharomyces cerevisiae and Saccharomyces carlsbergensis are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983)), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968); Holland et al. (1978)), including the promoter for 3 phosphoglycerate kinase (Hitzeman (1980)). Terminators may also be included, such as those derived from the enolase gene (Holland (1981)). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation

region which are operably linked may be such that they are not naturally associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985, all of which
5 are assigned to the herein assignee, and are hereby incorporated herein by reference.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture
10 Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers
15 (1978)), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression
20 may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.
25

III.B. Transformations

Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and
30 transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. For example, transformation of the E. coli host cells with lambda-gt11 containing BB-NANBV sequences is discussed in
35

the Example section, infra. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972); Maniatis (1982)). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978). Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), or the various known modifications thereof.

10 III.C. Vector Construction

Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 microgram of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 microliters buffer solution by incubation of 1-2 hr at 37° C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560.

25 Sticky ended cleavage fragments may be blunt ended using E. coli DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

30 Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often
35

treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

5

Ligation mixtures are transformed into suitable cloning hosts, such as E. coli, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

10

III.D. Construction of Desired DNA Sequences

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984). If desired the synthetic strands may be labeled with ³²P by treatment with polynucleotide kinase in the presence of ³²P-ATP, using standard conditions for the reaction.

15

DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase using, as a primer, a synthetic oligonucleotide complementary to the portion of the DNA to be modified, and having the desired modification included in its own sequence. The resulting double stranded DNA is transformed into a phage supporting host bacterium.

20

Cultures of the transformed bacteria, which contain replications of each strand of the phage, are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and conditions which permit hybridization

25

30

35

with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

5 III.E. Hybridization with Probe

DNA libraries may be probed using the procedure of Grunstein and Hogness (1975). Briefly, in this procedure, the DNA to be probed is immobilized on nitro-cellulose filters, denatured, and prehybridized with a buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinyl pyrrolidone, and Ficoll, 50 mM Na Phosphate (pH 6.5), 0.1% SDS, and 100 micrograms/ml carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42°C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'-³²P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

30 III.F. Verification of Construction and Sequencing

For routine vector constructions, ligation mixtures are transformed into E. coli strain HB101 or

other suitable host, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared according to the method of Clewell et al. (1969), usually following chloramphenicol amplification (Clewell (1972)). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing. Sequencing may be by the dideoxy method of Sanger et al. (1977) as further described by Messing et al. (1981), or by the method of Maxam et al. (1980). Problems with band compression, which are sometimes observed in GC rich regions, were overcome by use of T-deazoguanosine according to Barr et al. (1986).

15 III.G. Enzyme Linked Immunosorbent Assay

The enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added.

After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

5 IV. Examples

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous 10 embodiments within the scope of the claims will be apparent to those of ordinary skill in the art. The procedures set forth, for example, in Sections IV.A. may, if desired, be repeated but need not be, as techniques are available for construction of the desired nucleotide sequences based 15 on the information provided by the invention. Expression is exemplified in E. coli; however, other systems are available as set forth more fully in Section III.A. Additional epitopes derived from the genomic structure may also be produced, and used to generate antibodies as set 20 forth below.

IV.A. Preparation, Isolation and Sequencing of HCV cDNA

IV.A.1. Preparation of HCV cDNA

25 The source of NANB agent was a plasma pool derived from a chimpanzee with chronic NANBH. The chimpanzee had been experimentally infected with blood from another chimpanzee with chronic NANBH resulting from infection with HCV in a contaminated batch of factor 8 30 concentrate derived from pooled human sera. The chimpanzee plasma pool was made by combining many individual plasma samples containing high levels of alanine aminotransferase activity; this activity results from hepatic injury due to the HCV infection. Since 1 ml 35 of a 10^{-6} dilution of this pooled serum given i.v. caused

NANBH in another chimpanzee, its CID was at least 10^6 /ml, i.e., it had a high infectious virus titer.

A cDNA library from the high titer plasma pool was generated as follows. First, viral particles were isolated from the plasma; a 90 ml aliquot was diluted with 310 ml of a solution containing 50 mM Tris-HCl, pH 8.0, 1mM EDTA, 100 mM NaCl. Debris was removed by centrifugation for 20 min at 15,000 x g at 20°C. Viral particles in the resulting supernatant were then pelleted by centrifugation in a Beckman SW28 rotor at 28,000 rpm for 5 hours at 20°C. To release the viral genome, the particles were disrupted by suspending the pellets in 15 ml solution containing 1% sodium dodecyl sulfate (SDS), 10 mM EDTA, 10 mM Tris-HCl, pH 7.5, also containing 2 mg/ml proteinase k, followed by incubation at 45°C for 90 min. Nucleic acids were isolated by adding 0.8 micrograms MS2 bacteriophage RNA as carrier, and extracting the mixture four times with a 1:1 mixture of phenol:chloroform (phenol saturated with 0.5M Tris-HCl, pH 7.5, 0.1% (v/v) beta-mercaptoethanol, 0.1% (w/v) hydroxyquinolone, followed by extraction two times with chloroform. The aqueous phase was concentrated with 1-butanol prior to precipitation with 2.5 volumes absolute ethanol overnight at -20°C. Nucleic acid was recovered by centrifugation in a Beckman SW41 rotor at 40,000 rpm for 90 min at 4°C, and dissolved in water that had been treated with 0.05% (v/v) diethylpyrocarbonate and autoclaved.

Nucleic acid obtained by the above procedure (<2 micrograms) was denatured with 17.5 mM CH₃HgOH; cDNA was synthesized using this denatured nucleic acid as template, and was cloned into the EcoRI site of phage lambda-gt11 using methods described by Huynh (1985), except that random primers replaced oligo(dT) 12-18 during the synthesis of the first cDNA strand by reverse transcriptase (Taylor et al. (1976)). The resulting

double stranded cDNAs were fractionated according to size on a Sepharose CL-4B column; eluted material of approximate mean size 400, 300, 200, and 100 base-pairs were pooled into cDNA pools 1, 2, 3, and 4, respectively.

- 5 The lambda-gt11 cDNA library was generated from the cDNA in pool 3.

The lambda-gt11 cDNA library generated from pool 3 was screened for epitopes that could bind specifically with serum derived from a patient who had previously experienced NANBH. About 10^6 phage were screened with patient sera using the methods of Huynh et al. (1985), except that bound human antibody was detected with sheep anti-human Ig antisera that had been radio-labeled with ^{125}I . Five positive phages were identified and purified. 10
15 The five positive phages were then tested for specificity of binding to sera from 8 different humans previously infected with the NANBH agent, using the same method. Four of the phage encoded a polypeptide that reacted immunologically with only one human serum, i.e., the one 20 that was used for primary screening of the phage library. The fifth phage (5-1-1) encoded a polypeptide that reacted immunologically with 5 of 8 of the sera tested. Moreover, this polypeptide did not react immunologically with sera from 7 normal blood donors. Therefore, it appears that 25 clone 5-1-1 encodes a polypeptide which is specifically recognized immunologically by sera from NANB patients.

IV.A.2. Sequences of the HCV cDNA in Recombinant Phage 5-1-1, and of the Polypeptide Encoded Within the Sequence.

30 The cDNA in recombinant phage 5-1-1 was sequenced by the method of Sanger et al. (1977). Essentially, the cDNA was excised with EcoRI, isolated by size fractionation using gel electrophoresis. The EcoRI restriction fragments were subcloned into the M13 vectors, 35 mp18 and mp19 (Messing (1983)) and sequenced using the

dideoxychain termination method of Sanger et al. (1977).
The sequence obtained is shown in Fig. 1.

The polypeptide encoded in Fig. 1 that is encoded in the HCV cDNA is in the same translational frame 5 as the N-terminal beta-galactosidase moiety to which it is fused. As shown in Section IV.A., the translational open reading frame (ORF) of 5-1-1 encodes epitope(s) specifically recognized by sera from patients and chimpanzees with NANBH infections.

10

IV.A.3. Isolation of Overlapping HCV cDNA to cDNA in Clone 5-1-1.

Overlapping HCV cDNA to the cDNA in clone 5-1-1 was obtained by screening the same lambda-gt11 library, 15 created as described in Section IV.A.1., with a synthetic polynucleotide derived from the sequence of the HCV cDNA in clones 5-1-1, as shown in Fig. 1. The sequence of the polynucleotide used for screening was:

20

5'-TCC CTT GCT CGA TGT ACG GTA AGT GCT GAG AGC
ACT CTT CCA TCT CAT CGA ACT CTC GGT AGA GGA CTT CCC TGT
CAG GT-3'.

The lambda-gt11 library was screened with this probe, 25 using the method described in Huynh (1985). Approximately 1 in 50,000 clones hybridized with the probe. Three clones which contained cDNAs which hybridized with the synthetic probe have been numbered 81, 1-2, and 91.

30 IV.A.4. Nucleotide Sequences of Overlapping HCV cDNAs to cDNA in Clone 5-1-1.

The nucleotide sequences of the three cDNAs in clones 81, 1-2, and 91 were determined essentially as in Section IV.A.2. The sequences of these clones relative to 35 the HCV cDNA sequence in phage 5-1-1 is shown in Fig. 2,

which shows the strand encoding the detected HCV epitope, and where the homologies in the nucleotide sequences are indicated by vertical lines between the sequences.

The sequences of the cloned HCV cDNAs are highly
5 homologous in the overlapping regions (see Fig. 2). However, there are differences in two regions. Nucleotide 67 in clone 1-2 is a thymidine, whereas the other three clones contain a cytidine residue in this position. It
10 should be noted, however, that the same amino acid is encoded when either C or T occupies this position.

The second difference is that clone 5-1-1 contains 28 base pairs which are not present in the other three clones. These base pairs occur at the start of the
15 cDNA sequence in 5-1-1, and are indicated by small letters. Based on radioimmunoassay data, which is discussed infra in Section IV.D., it is possible that an HCV epitope may be encoded in this 28 bp region.

The absence of the 28 base pairs of 5-1-1 from
20 clones 81, 1-2, and 91 may mean that the cDNA in these clones were derived from defective HCV genomes; alternatively, the 28 bp region could be a terminal artifact in clone 5-1-1.

The sequences of small letters in the nucleotide
25 sequence of clones 81 and 91 simply indicate that these sequences have not been found in other cDNAs because cDNAs overlapping these regions were not yet isolated.

A composite HCV cDNA sequence derived from overlapping cDNAs in clones 5-1-1, 81, 1-2 and 91 is shown in
30 Fig. 3. However, in this figure the unique 28 base pairs of clone 5-1-1 are omitted. The figure also shows the sequence of the polypeptide encoded within the ORF of the composite HCV cDNA.

IV.A.5. Isolation of Overlapping HCV cDNAs to cDNA in Clone 81.

The isolation of HCV cDNA sequences upstream of, and which overlap those in clone 81 cDNA was accomplished 5 as follows. The lambda-gt11 cDNA library prepared as described in Section IV.A.1. was screened by hybridization with a synthetic polynucleotide probe which was homologous to a 5' terminal sequence of clone 81. The sequence of clone 81 is presented in Fig. 4. The sequence of the 10 synthetic polynucleotide used for screening was:

5' CTG TCA GGT ATG ATT GCC GGC TTC CCG GAC 3'.

The methods were essentially as described in Huynh (1985), 15 except that the library filters were given two washes under stringent conditions, i.e., the washes were in 5 x SSC, 0.1% SDS at 55°C for 30 minutes each. Approximately 1 in 50,000 clones hybridized with the probe. A positive recombinant phage which contained cDNA which hybridized 20 with the sequence was isolated and purified. This phage has been numbered clone 36.

Downstream cDNA sequences, which overlaps the carboxyl-end sequences in clone 81 cDNA were isolated using a procedure similar to that for the isolation of 25 upstream cDNA sequences, except that a synthetic oligonucleotide probe was prepared which is homologous to a 3' terminal sequence of clone 81. The sequence of the synthetic polynucleotide used for screening was:

30 5' TTT GGC TAG TGG TTA GTG GGC TGG TGA CAG 3'

A positive recombinant phage, which contained cDNA which hybridized with this latter sequence was isolated and purified, and has been numbered clone 32.

IV.A.6. Nucleotide Sequence of HCV cDNA in Clone 36.

The nucleotide sequence of the cDNA in clone 36 was determined essentially as described in Section IV.A.2. The double-stranded sequence of this cDNA, its region of 5 overlap with the HCV cDNA in clone 81, and the polypeptide encoded by the ORF are shown in Fig. 5.

The ORF in clone 36 is in the same translational frame as the HCV antigen encoded in clone 81. Thus, in combination, the ORFs in clones 36 and 81 encode a 10 polypeptide that represents part of a large HCV antigen. The sequence of this putative HCV polypeptide and the double stranded DNA sequence encoding it, which is derived from the combined ORFs of the HCV cDNAs of clones 36 and 81, is shown in Fig. 6.

15

IV.A.7 Nucleotide Sequences of HCV cDNA in Clone 32

The nucleotide sequence of the cDNA in clone 32 was determined essentially as was that described in Section IV.A.2 for the sequence of clone 5-1-1. The sequence 20 data indicated that the cDNA in clone 32 recombinant phage was derived from two different sources. One fragment of the cDNA was comprised of 418 nucleotides derived from the HCV genome; the other fragment was comprised of 172 nucleotides derived from the bacteriophage MS2 genome, 25 which had been used as a carrier during the preparation of the lambda gt11 plasma cDNA library.

The sequence of the cDNA in clone 32 corresponding to that of the HCV genome is shown in Fig. 7. The region of the sequences that overlaps that of clone 81, 30 and the polypeptide encoded by the ORF are also indicated in the figure. This sequence contains one continuous ORF that is in the same translational frame as the HCV antigen encoded by clone 81.

IV.A.8 Isolation of Overlapping HCV cDNA to cDNA in Clone 36

The isolation of HCV cDNA sequences upstream of, and which overlap those in clone 36 cDNA was accomplished 5 as described in Section IV.A.5, for those which overlap clone 81 cDNA, except that the synthetic polynucleotide was based on the 5'-region of clone 36. The sequence of the synthetic polynucleotide used for screening was:

10 5' AAG CCA CCG TGT GCG CTA GGG CTC AAG CCC 3'

Approximately 1 in 50,000 clones hybridized with the probe. The isolated, purified clone of recombinant phage which contained cDNA which hybridized to this sequence was 15 named clone 35.

IV.A.9 Nucleotide Sequence of HCV cDNA in Clone 35

The nucleotide sequence of the cDNA in clone 35 was determined essentially as described in Section IV.A.2. 20 The sequence, its region of overlap with that of the cDNA in clone 36, and the putative polypeptide encoded therein, are shown in Fig. 8.

Clone 35 apparently contains a single, continuous ORF that encodes a polypeptide in the same 25 translational frame as that encoded by clone 36, clone 81, and clone 32. Fig. 9 shows the sequence of the long continuous ORF that extends through clones 35, 36, 81, and 32, along with the putative HCV polypeptide encoded therein. This combined sequence has been confirmed using 30 other independent cDNA clones derived from the same lambda gt11 cDNA library.

IV.A.10. Isolation of Overlapping HCV cDNA to cDNA in Clone 35

The isolation of HCV cDNA sequences upstream of, and which overlap those in clone 35 cDNA was accomplished 5 as described in Section IV.A.8, for those which overlap clone 36 cDNA, except that the synthetic polynucleotide was based on the 5'-region of clone 35. The sequence of the synthetic polynucleotide used for screening was:

10 5' CAG GAT GCT GTC TCC CGC ACT CAA CGT 3'

Approximately 1 in 50,000 clones hybridized with the probe. The isolated, purified clone of recombinant phage which contained cDNA which hybridized to this sequence was 15 named clone 37b.

IV.A.11. Nucleotide Sequence of HCV in Clone 37b

The nucleotide sequence of the cDNA in clone 37b was determined essentially as described in Section IV.A.2. 20 The sequence, its region of overlap with that of the cDNA in clone 35, and the putative polypeptide encoded therein, are shown in Fig. 10.

The 5'-terminal nucleotide of clone 35 is a T, whereas the corresponding nucleotide in clone 37b is an A. 25 The cDNAs from three other independent clones which were isolated during the procedure in which clone 37b was isolated, described in Section IV.A.10, have also been sequenced. The cDNAs from these clones also contain an A in this position. Thus, the 5'-terminal T in clone 35 may 30 be an artefact of the cloning procedure. It is known that artefacts often arise at the 5'-termini of cDNA molecules.

Clone 37b apparently contains one continuous ORF which encodes a polypeptide which is a continuation of the polypeptide encoded in the ORF which extends through the 35 overlapping clones 35, 36, 81 and 32.

IV.A.12 Isolation of Overlapping HCV cDNA to cDNA in Clone 32

The isolation of HCV cDNA sequences downstream of clone 32 was accomplished as follows. First, clone cla 5 was isolated utilizing a synthetic hybridization probe which was based on the nucleotide sequence of the HCV cDNA sequence in clone 32. The method was essentially that described in Section IV.A.5, except that the sequence of the synthetic probe was:

10

5' AGT GCA GTG GAT GAA CCG GCT GAT AGC CTT 3'.

Utilizing the nucleotide sequence from clone cla, another synthetic nucleotide was synthesized which had the 15 sequence:

5' TCC TGA GGC GAC TGC ACC AGT GGA TAA GCT 3'.

Screening of the lambda gt11 library using the clone cla 20 derived sequence as probe yielded approximately 1 in 50,000 positive colonies. An isolated, purified clone which hybridized with this probe was named clone 33b.

IV.A.13 Nucleotide Sequence of HCV cDNA in Clone 33b

25 The nucleotide sequence of the cDNA in clone 33b was determined essentially as described in Section IV.A.2. The sequence, its region of overlap with that of the cDNA in clone 32, and the putative polypeptide encoded therein, are shown in Fig. 11.

30 Clone 33b apparently contains one continuous ORF which is an extension of the ORFs in overlapping clones 37b, 35, 36, 81 and 32. The polypeptide encoded in clone 33b is in the same translational frame as that encoded in the extended ORF of these overlapping clones.

35

IV.A.14 Isolation of Overlapping HCV cDNAs to cDNA Clone 37b and to cDNA in Clone 33b

In order to isolate HCV cDNAs which overlap the cDNAs in clone 37b and in clone 33b, the following

5 synthetic oligonucleotide probes, which were derived from the cDNAs in those clones, were used to screen the lambda gt11 library, using essentially the method described in Section IV.A.3. The probes used were:

10 5' CAG GAT GCT GTC TCC CGC ACT CAA CGT C 3'

and

5' TCC TGA GGC GAC TGC ACC AGT GGA TAA GCT 3'

15

to detect colonies containing HCV cDNA sequences which overlap those in clones 37b and 33b, respectively. Approximately 1 in 50,000 colonies were detected with each probe. A clone which contained cDNA which was upstream 20 of, and which overlapped the cDNA in clone 37b, was named clone 40b. A clone which contained cDNA which was downstream of, and which overlapped the cDNA in clone 33b was named clone 25c.

25

IV.A.15 Nucleotide Sequences of HCV cDNA in clone 40b and in clone 25c

The nucleotide sequences of the cDNAs in clone 40b and in clone 25c were determined essentially as 30 described in Section IV.A.2. The sequences of 40b and 25c, their regions of overlap with the cDNAs in clones 37b and 33b, and the putative polypeptides encoded therein, are shown in Fig. 12 (clone 40b) and Fig. 13 (clone 25c).

The 5'-terminal nucleotide of clone 40b is a G. 35 However, the cDNAs from five other independent clones

which were isolated during the procedure in which clone 40b was isolated, described in Section IV.A.14, have also been sequenced. The cDNAs from these clones also contain a T in this position. Thus, the G may represent a cloning artifact (see the discussion in Section IV.A.11).
5

The 5'-terminus of clone 25c is ACT, but the sequence of this region in clone cla (sequence not shown), and in clone 33b is TCA. This difference may also represent a cloning artifact, as may the 28 extra 5'-
10 terminal nucleotides in clone 5-1-1.

Clones 40b and 25c each apparently contain an ORF which is an extension of the continuous ORF in the previously sequenced clones. The nucleotide sequence of the ORF extending through clones 40b, 37b, 35, 36, 81, 32,
15 33b, and 25c, and the amino acid sequence of the putative polypeptide encoded therein, are shown in Fig. 14. In the figure, the potential artifacts have been omitted from the sequence, and instead, the corresponding sequences in non-5'-terminal regions of multiple overlapping clones are
20 shown.

IV.A.16. Preparation of a Composite HCV cDNA from the cDNAs in Clones 36, 81, and 32

The composite HCV cDNA, C100, was constructed as follows. First the cDNAs from the clones 36, 81, and 32 were excised with EcoRI. The EcoRI fragment of cDNA from each clone was cloned individually into the EcoRI site of the vector pGEM3-blue (Promega Biotec). The resulting recombinant vectors which contained the cDNAs from clones 36, 81, and 32 were named pGEM3-blue/36, pGEM3-blue/81,
25 and pGEM3-blue/32, respectively. The appropriately oriented recombinant of pGEM3-blue/81 was digested with NaeI and NarI, and the large (~2850bp) fragment was purified and ligated with the small (~570bp) NaeI/NarI purified restriction fragment from pGEM3-blue/36. This
30
35

composite of the cDNAs from clones 36 and 81 was used to generate another pGEM3-blue vector containing the continuous HCV ORF contained within the overlapping cDNA within these clones. This new plasmid was then digested

5 with PvuII and EcoRI to release a fragment of approximately 680bp, which was then ligated with the small (580bp) PvuII/EcoRI fragment isolated from the appropriately oriented pGEM3-blue/32 plasmid, and the composite cDNA from clones 36, 81, and 32 was ligated into

10 the EcoRI linearized vector pSODcf1, which is described in Section IV.B.1, and which was used to express clone 5-1-1 in bacteria. Recombinants containing the ~1270bp EcoRI fragment of composite HCV cDNA (C100) were selected, and the cDNA from the plasmids was excised with EcoRI and

15 purified.

IV.A.17. Isolation and Nucleotide Sequences of HCV cDNAs in Clones 14i, 11b, 7f, 7e, 8h, 33c, 14c, 8f, 33f, 33g, and 39c

20 The HCV cDNAs in clones 14i, 11b, 7f, 7e, 8h, 33c, 14c, 8f, 33f, 33g, and 39c were isolated by the technique of isolating overlapping cDNA fragments from the lambda gt11 library of HCV cDNAs described in Section IV.A.1.. The technique used was essentially as described

25 in Section IV.A.3., except that the probes used were designed from the nucleotide sequence of the last isolated clones from the 5' and the 3' end of the combined HCV sequences. The frequency of clones which hybridized with the probes described below was approximately 1 in 50,000

30 in each case.

The nucleotide sequences of the HCV cDNAs in clones 14i, 7f, 7e, 8h, 33c, 14c, 8f, 33f, 33g, and 39c were determined essentially as described in Section IV.A.2., except that the cDNA excised from these phages

35 were substituted for the cDNA isolated from clone 5-1-1.

Clone 33c was isolated using a hybridization probe based on the sequence of nucleotides in clone 40b. The nucleotide sequence of clone 40b is presented in Fig. 12. The nucleotide sequence of the probe used to isolate

5 33c was:

5' ATC AGG ACC GGG GTG AGA ACA ATT ACC ACT 3'

The sequence of the HCV cDNA in clone 33c, and the overlap
10 with that in clone 40b, is shown in Fig. 15, which also shows the amino acids encoded therein.

Clone 8h was isolated using a probe based on the sequence of nucleotides in clone 33c. The nucleotide sequence of the probe was

15

5' AGA GAC AAC CAT GAG GTC CCC GGT GTT C 3'.

The sequence of the HCV cDNA in clone 8h, and the overlap with that in clone 33c, and the amino acids encoded
20 therein, are shown in Fig. 16.

Clone 7e was isolated using a probe based on the sequence of nucleotides in clone 8h. The nucleotide sequence of the probe was

25

5' TCG GAC CTT TAC CTG GTC ACG AGG CAC 3'.

The sequence of HCV cDNA in clone 7e, the overlap with clone 8h, and the amino acids encoded therein, are shown in Fig. 17.

30

Clone 14c was isolated with a probe based on the sequence of nucleotides in clone 25c. The sequence of clone 25c is shown in Fig. 13. The probe in the isolation of clone 14c had the sequence

35

5' ACC TTC CCC ATT AAT GCC TAC ACC ACG GGC 3'.

The sequence of HCV cDNA in clone 14c, its overlap with that in clone 25c, and the amino acids encoded therein are shown in Fig. 18.

Clone 8f was isolated using a probe based on the
5 sequence of nucleotides in clone 14c. The nucleotide sequence of the probe was

5' TCC ATC TCT CAA GGC AAC TTG CAC CGC TAA 3'.

10 The sequence of HCV cDNA in clone 8f, its overlap with that in clone 14c, and the amino acids encoded therein are shown in Fig. 19.

Clone 33f was isolated using a probe based on the nucleotide sequence present in clone 8f. The
15 nucleotide sequence of the probe was

5' TCC ATG GCT GTC CGC TTC CAC CTC CAA AGT 3'.

The sequence of HCV cDNA in clone 33f, its overlap with
20 that in clone 8f, and the amino acids encoded therein are shown in Fig. 20.

Clone 33g was isolated using a probe based on the sequence of nucleotides in clone 33f. The nucleotide sequence of the probe was
25

5' GCG ACA ATA CGA CAA CAT CCT CTG AGC CCG 3'.

The sequence of HCV cDNA in clone 33g, its overlap with that in clone 33f, and the amino acids encoded therein are
30 shown in Fig. 21.

Clone 7f was isolated using a probe based on the sequence of nucleotides in clone 7e. The nucleotide sequence of the probe was

35 5' AGC AGA CAA GGG GCC TCC TAG GGT GCA TAA T 3'.

The sequence of HCV cDNA in clone 7f, its overlap with clone 7e, and the amino acids encoded therein are shown in Fig. 22.

Clone 11b was isolated using a probe based on 5 the sequence of clone 7f. The nucleotide sequence of the probe was

5' CAC CTA TGT TTA TAA CCA TCT CAC TCC TCT 3'.

10 The sequence of HCV cDNA in clone 11b, its overlap with clone 7f, and the amino acids encoded therein are shown in Fig. 23.

Clone 14i was isolated using a probe based on 15 the sequence of nucleotides in clone 11b. The nucleotide sequence of the probe was

5' CTC TGT CAC CAT ATT ACA AGC GCT ATA TCA 3'.

20 The sequence of HCV cDNA in clone 14i, its overlap with 24. 11b, and the amino acids encoded therein are shown in Fig. 24.

Clone 39c was isolated using a probe based on 25 the sequence of nucleotides in clone 33g. The nucleotide sequence of the probe was

5' CTC GTT GCT ACG TCA CCA CAA TTT GGT GTA 3'

30 The sequence of HCV cDNA in clone 39c, its overlap with clone 33g, and the amino acids encoded therein are shown in Fig. 25.

IV.A.18. The Composite HCV cDNA Sequence Derived from Isolated Clones Containing HCV cDNA

35 The HCV cDNA sequences in the isolated clones described supra have been aligned to create a composite

HCV cDNA sequence. The isolated clones, aligned in the 5' to 3' direction are: 14i, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, and 39c.

A composite HCV cDNA sequence derived from the 5 isolated clones, and the amino acids encoded therein, is shown in Fig. 26.

In creating the composite sequence the following sequence heterogeneities have been considered. Clone 33c contains an HCV cDNA of 800 base pairs, which overlaps the 10 cDNAs in clones 40b and 37c. In clone 33c, as well as in 5 other overlapping clones, nucleotide #789 is a G. However, in clone 37b (see Section IV.A.11), the corresponding nucleotide is an A. This sequence difference creates an apparent heterogeneity in the amino acids 15 encoded therein, which would be either CYS or TYR, for G or A, respectively. This heterogeneity may have important ramifications in terms of protein folding.

Nucleotide residue #2 in clone 8h HCV cDNA is a T. However, as shown infra, the corresponding residue in 20 clone 7e is an A; moreover, an A in this position is also found in 3 other isolated overlapping clones. Thus, the T residue in clone 8h may represent a cloning artifact. Therefore, in Fig. 26, the residue in this position is designated as an A.

25 The 3'-terminal nucleotide in clone 8f HCV cDNA is a G. However, the corresponding residue in clone 33f, and in 2 other overlapping clones is a T. Therefore, in Fig. 26, the residue in this position is designated as a T.

30 The 3'-terminal sequence in clone 33f HCV cDNA is TTGC. However, the corresponding sequence in clone 33g and in 2 other overlapping clones is ATTC. Therefore, in Fig. 26, the corresponding region is represented as ATTC.

35 Nucleotide residue #4 in clone 33g HCV cDNA is a T. However, in clone 33f and in 2 other overlapping

clones the corresponding residue is an A. Therefore, in Fig. 26, the corresponding residue is designated as an A.

The 3'-terminus of clone 14i is an AA, whereas the corresponding dinucleotide in clone 11b, and in three 5 other clones, is TA. Therefore, in Fig. 26, the TA residue is depicted.

The resolution of other sequence heterogeneities is discussed supra.

An examination of the composite HCV cDNA 10 indicates that it contains one large ORF. This suggests that the viral genome is translated into a large polypeptide which is processed concomitant with, or subsequent to translation.

15 IV.A.19. Isolation and Nucleotide Sequences of HCV cDNAs in Clones 12f, 35f, 19g, 26g, and 15e

The HCV cDNAs in clones 12f, 35f, 19g, 26g, and 15e were isolated essentially by the technique described in Section IV.A.17, except that the probes were as 20 indicated below. The frequency of clones which hybridized with the probes was approximately 1 in 50,000 in each case. The nucleotide sequences of the HCV cDNAs in these clones were determined essentially as described in Section IV.A.2., except that the cDNA from the indicated clones 25 were substituted for the cDNA isolated from clone 5-1-1.

The isolation of clone 12f, which contains cDNA upstream of the HCV cDNA in Fig. 26, was accomplished using a hybridization probe based on the sequence of nucleotides in clone 14i. The nucleotide sequence of the 30 probe was

5' TGC TTG TGG ATG ATG CTA CTC ATA TCC CAA 3'.

The HCV cDNA sequence of clone 12f, its overlap with clone 14i, and the amino acids encoded therein are shown in Fig. 27.

5 The isolation of clone 35f, which contains cDNA downstream of the HCV cDNA in Fig. 26, was accomplished using a hybridization probe based on the sequence of nucleotides in clone 39c. The nucleotide sequence of the probe was

10 5' AGC AGC GGC GTC AAA AGT GAA GGC TAA CTT 3'.

The sequence of clone 35f, its overlap with the sequence in clone 39c, and the amino acids encoded therein are shown in Fig. 28.

15 The isolation of clone 19g was accomplished using a hybridization probe based on the 3' sequence of clone 35f. The nucleotide sequence of the probe was

5' TTC TCG TAT GAT ACC CGC TGC TTT GAC TCC 3'.

20 The HCV cDNA sequence of clone 19g, its overlap with the sequence in clone 35f, and the amino acids encoded therein are shown in Fig. 29.

25 The isolation of clone 26g was accomplished using a hybridization probe based on the 3' sequence of clone 19g. The nucleotide sequence of the probe was

5' TGT GTG GCG ACG ACT TAG TCG TTA TCT GTG 3'.

30 The HCV cDNA sequence of clone 26g, its overlap with the sequence in clone 19g, and the amino acids encoded therein are shown in Fig. 30.

Clone 15e was isolated using a hybridization probe based on the 3' sequence of clone 26 g. The
35 nucleotide sequence of the probe was

5' CAC ACT CCA GTC AAT TCC TGG CTA GGC AAC 3'.

The HCV cDNA sequence of clone 15e, its overlap with the sequence in clone 26g, and the amino acids encoded therein
5 are shown in Fig. 31.

The clones described in this Section have been deposited with the ATCC under the terms and conditions described in Section II.A., and have been assigned the following Accession Numbers.

10

	<u>lambda-gt11</u>	<u>ATCC No.</u>	<u>Deposit Date</u>
	clone 12f	40514	10 November 1988
	clone 35f	40511	10 November 1988
	clone 15e	40513	10 November 1988
15	clone k9-1	40512	10 November 1988

The HCV cDNA sequences in the isolated clones described supra. have been aligned to create a composite
20 HCV cDNA sequence. The isolated clones, aligned in the 5' to 3' direction are: 12f, 14i, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f 33f, 33g, 39c, 35f, 19g, 26g, and 15e.

A composite HCV cDNA sequence derived from the
25 isolated clones, and the amino acids encoded therein, is shown in Fig. 32.

IV.A.20. Alternative Method of Isolating cDNA Sequences Upstream of the HCV cDNA Sequence in Clone 12f

30 Based on the most 5' HCV sequence in Fig. 32, which is derived from the HCV cDNA in clone 12f, small synthetic oligonucleotide primers of reverse transcriptase are synthesized and used to bind to the corresponding sequence in HCV genomic RNA, to prime reverse
35 transcription of the upstream sequences. The primer

sequences are proximal to the known 5'-terminal sequence of clone 12f, but sufficiently downstream to allow the design of probe sequences upstream of the primer sequences. Known standard methods of priming and cloning
5 are used. The resulting cDNA libraries are screened with sequences upstream of the priming sites (as deduced from the elucidated sequence in clone 12f). The HCV genomic RNA is obtained from either plasma or liver samples from chimpanzees with NANBH, or from analogous samples from
10 humans with NANBH.

IV.A.21. Alternative Method Utilizing Tailing to Isolate Sequences from the 5'-Terminal Region of the HCV Genome

In order to isolate the extreme 5'-terminal
15 sequences of the HCV RNA genome, the cDNA product of the first round of reverse transcription, which is duplexed with the template RNA, is tailed with oligo C. This is accomplished by incubating the product with terminal transferase in the presence of CTP. The second round of
20 cDNA synthesis, which yields the complement of the first strand of cDNA, is accomplished utilizing oligo G as a primer for the reverse transcriptase reaction. The sources of genomic HCV RNA are as described in Section IV.A.20. The methods for tailing with terminal
25 transferase, and for the reverse transcriptase reactions are as in Maniatis et al. (1982). The cDNA products are then cloned, screened, and sequenced.

IV.A.22. Alternative Method Utilizing Tailing to Isolate Sequences from the 3'-Terminal Region of the HCV Genome

This method is based on previously used methods for cloning cDNAs of Flavivirus RNA. In this method, the RNA is subjected to denaturing conditions to remove secondary structures at the 3'-terminus, and is then
35 tailed with Poly A polymerase using rATP as a substrate.

Reverse transcription of the poly A tailed RNA is catalyzed by reverse transcriptase, utilizing oligo dT as a primer. The second strands of cDNA are synthesized, the cDNA products are cloned, screened, and sequenced.

5

IV.A.23 Creation of Lambda-gt11 HCV cDNA Libraries Containing Larger cDNA Inserts

The method used to create and screen the Lambda gt11 libraries are essentially as described in Section 10 IV.A.1., except that the library is generated from a pool of larger size cDNAs eluted from the Sepharose CL-4B column.

15 IV.A.24. Creation of HCV cDNA Libraries Using Synthetic Oligomers as Primers

New HCV cDNA libraries have been prepared from the RNA derived from the infectious chimpanzee plasma pool described in Section IV.A.1., and from the poly A⁺ RNA fraction derived from the liver of this infected animal. 20 The cDNA was constructed essentially as described by Gubler and Hoffman (1983), except that the primers for the first cDNA strand synthesis were two synthetic oligomers based on the sequence of the HCV genome described supra. Primers based on the sequence of clone 11 b and 7e were, 25 respectively,

5' CTG GCT TGA AGA ATC 3'

and

30

5' AGT TAG GCT GGT GAT TAT GC 3'.

The resulting cDNAs were cloned into lambda bacteriophage vectors, and screened with various other synthetic 35

oligomers, whose sequences were based on the HCV sequence in Fig. 32.

5 IV.B. Expression of Polypeptides Encoded Within HCV cDNAs
and Identification of the Expressed Products as HCV
Induced Antigens.

IV.B.1. Expression of the Polypeptide Encoded in Clone 5-1-1.

10 The HCV polypeptide encoded within clone 5-1-1
(see Section IV.A.2., supra) was expressed as a fusion
polypeptide with superoxide dismutase (SOD). This was
accomplished by subcloning the clone 5-1-1 cDNA insert
into the expression vector pSODcfl (Steimer et al. (1986))
15 as follows.

First, DNA isolated from pSODcf1 was treated with BamHI and EcoRI, and the following linker was ligated into the linear DNA created by the restriction enzymes:

20 5' GAT CCT GGA ATT CTG ATA A 3'
 3' GA CCT TAA GAC TAT TTT AA 5'

After cloning, the plasmid containing the insert was isolated.

25 Plasmid containing the insert was restricted
with EcoRI. The HCV cDNA insert in clone 5-1-1 was
excised with EcoRI, and ligated into this EcoRI linearized
plasmid DNA. The DNA mixture was used to transform E.
coli strain D1210 (Sadler et al. (1980)). Recombinants
30 with the 5-1-1 cDNA in the correct orientation for
expression of the ORF shown in Fig. 1 were identified by
restriction mapping and nucleotide sequencing.

Recombinant bacteria from one clone were induced
to express the SOD-NANB₅₋₁₋₁ polypeptide by growing the
bacteria in the presence of IPTG.

IV.B.2. Expression of the Polypeptide Encoded in Clone 81.

The HCV cDNA contained within clone 81 was expressed as a SOD-NANB₈₁ fusion polypeptide. The method 5 for preparing the vector encoding this fusion polypeptide was analogous to that used for the creation of the vector encoding SOD-NANB₅₋₁₋₁, except that the source of the HCV cDNA was clone 81, which was isolated as described in Section IV.A.3, and for which the cDNA sequence was 10 determined as described in Section IV.A.4. The nucleotide sequence of the HCV cDNA in clone 81, and the putative amino acid sequence of the polypeptide encoded therein are shown in Fig. 4.

The HCV cDNA insert in clone 81 was excised with 15 EcoRI, and ligated into the pSODcfl which contained the linker (see IV.B.1.) and which was linearized by treatment with EcoRI. The DNA mixture was used to transform E. coli strain D1210. Recombinants with the clone 81 HCV cDNA in the correct orientation for expression of the ORF shown in 20 Fig. 4 were identified by restriction mapping and nucleotide sequencing.

Recombinant bacteria from one clone were induced to express the SOD-NANB₈₁ polypeptide by growing the bacteria in the presence of IPTG.

25

IV.B.3. Identification of the Polypeptide Encoded Within Clone 5-1-1 as an HCV and NANBH Associated Antigen.

The polypeptide encoded within the HCV cDNA of clone 5-1-1 was identified as a NANBH associated antigen 30 by demonstrating that sera of chimpanzees and humans infected with NANBH reacted immunologically with the fusion polypeptide, SOD-NANB₅₋₁₋₁, which is comprised of superoxide dismutase at its N-terminus and the in-frame 5-1-1 antigen at its C-terminus. This was accomplished by 35 "Western" blotting (Towbin et al. (1979)) as follows.

A recombinant strain of bacteria transformed with an expression vector encoding the SOD-NANB₅₋₁₋₁ polypeptide, described in Section IV.B.I., was induced to express the fusion polypeptide by growth in the presence 5 of IPTG. Total bacterial lysate was subjected to electrophoresis through polyacrylamide gels in the presence of SDS according to Laemmli (1970). The separated polypeptides were transferred onto nitrocellulose filters (Towbin et al. (1979)). The filters were then cut into 10 thin strips, and the strips were incubated individually with the different chimpanzee and human sera. Bound antibodies were detected by further incubation with ¹²⁵I-labeled sheep anti-human Ig, as described in Section IV.A.1.

15 The characterization of the chimpanzee sera used for the Western blots and the results, shown in the photograph of the autoradiographed strips, are presented in Fig. 33. Nitrocellulose strips containing polypeptides were incubated with sera derived from chimpanzees at different times during acute NANBH (Hutchinson strain) infections (lanes 1-16), hepatitis A infections (lanes 17-24, and 26-33), and hepatitis B infections (lanes 34-44). Lanes 25 and 45 show positive controls in which the immunoblots were incubated with serum from the patient 20 used to identify the recombinant clone 5-1-1 in the original screening of the lambda-gt11 cDNA library (see 25 Section IV.A.1.).

The band visible in the control lanes, 25 and 45, in Fig. 23 reflects the binding of antibodies to the 30 NANB₅₋₁₋₁ moiety of the SOD fusion polypeptide. These antibodies do not exhibit binding to SOD alone, since this has also been included as a negative control in these samples, and would have appeared as a band migrating significantly faster than the SOD-NANB₅₋₁₋₁ fusion 35 polypeptide.

Lanes 1-16 of Fig. 33 show the binding of antibodies in sera samples of 4 chimpanzees; the sera were obtained just prior to infection with NANBH, and sequentially during acute infection. As seen from the 5 figure, whereas antibodies which reacted immunologically with the SOD-NANB₅₋₁₋₁ polypeptide were absent in sera samples obtained before administration of infectious HCV inoculum and during the early acute phase of infection, all 4 animals eventually induced circulating antibodies to 10 this polypeptide during the late part of, or following the acute phase. Additional bands observed on the immunoblots in the cases of chimps numbers 3 and 4 were due to background binding to host bacterial proteins.

In contrast to the results obtained with sera 15 from chimps infected with NANBH, the development of antibodies to the NANB₅₋₁₋₁ moiety of the fusion polypeptide was not observed in 4 chimpanzees infected with HAV or 3 chimpanzees infected with HBV. The only binding in these cases was background binding to the host bacterial 20 proteins, which also occurred in the HCV infected samples.

The characterization of the human sera used for the Western blots, and the results, which are shown in the photograph of the autoradiographed strips, are presented in Fig. 34. Nitrocellulose strips containing polypeptides 25 were incubated with sera derived from humans at different times during infections with NANBH (lanes 1-21), HAV (lanes 33-40), and HBV (lanes 41-49). Lanes 25 and 50 show positive controls in which the immunoblots were incubated with serum from patient used in the original 30 screening of the lambda-gt11 library, described supra. Lanes 22-24 and 26-32 show "non-infected" controls in which the sera was from "normal" blood donors.

As seen in Fig. 34, sera from nine NANBH patients, including the serum used for screening the 35 lambda-gt11 library, contained antibodies to the NANB₅₋₁₋₁

moiety of the fusion polypeptide. Sera from three patients with NANBH did not contain these antibodies. It is possible that the anti-NANB₅₋₁₋₁ antibodies will develop at a future date in these patients. It is also 5 possible that this lack of reaction resulted from a different NANBV agent being causative of the disease in the individuals from which the non-responding serum was taken.

Fig. 34 also shows that sera from many patients infected with HAV and HBV did not contain anti-NANB₅₋₁₋₁ 10 antibodies, and that these antibodies were also not present in the sera from "normal" controls. Although one HAV patient (lane 36) appears to contain anti-NANB₅₋₁₋₁ antibodies, it is possible that this patient had been previously infected with HCV, since the incidence of NANBH 15 is very high and since it is often subclinical.

These serological studies indicate that the cDNA in clone 5-1-1 encodes epitopes which are recognized specifically by sera from patients and animals infected with BB-NANBV. In addition, the cDNA does not appear to 20 be derived from the primate genome. A hybridization probe made from clone 5-1-1 or from clone 81 did not hybridize to "Southern" blots of control human and chimpanzee genomic DNA from uninfected individuals under conditions where unique, single-copy genes are detectable. These 25 probes also did not hybridize to Southern blots of control bovine genomic DNA.

IV.B.4. Expression of the Polypeptide Encoded in a Composite of the HCV cDNAs in Clones 36, 81 and 32

30 The HCV polypeptide which is encoded in the ORF which extends through clones 36, 81 and 32 was expressed as a fusion polypeptide with SOD. This was accomplished by inserting the composite cDNA, C100, into an expression cassette which contains the human superoxide dismutase 35 gene, inserting the expression cassette into a yeast

expression vector, and expressing the polypeptide in yeast.

An expression cassette containing the composite C100 cDNA derived from clones 36, 81, and 32, was 5 constructed by inserting the ~1270bp EcoRI fragment into the EcoRI site of the vector pS3-56 (also called pS356), yielding the plasmid pS3-56_{C100}. The construction of C100 is described in Section IV.A.16, supra.

The vector pS3-56, which is a pBR322 derivative, 10 contains an expression cassette which is comprised of the ADH2/GAPDH hybrid yeast promoter upstream of the human superoxide dismutase gene, and a downstream GAPDH transcription terminator. A similar cassette, which contains these control elements and the superoxide 15 dismutase gene has been described in Cousens et al. (1987), and in copending application EPO 196,056, published October 1, 1986, which is commonly owned by the herein assignee. The cassette in pS3-56, however, differs from that in Cousens et al. (1987) in that the 20 heterologous proinsulin gene and the immunoglobulin hinge are deleted, and in that the gln₁₅₄ of the superoxide dismutase is followed by an adaptor sequence which contains an EcoRI site. The sequence of the adaptor is:

25 5'-AAT TTG GGA ATT CCA TAA TGA G -3'
 AC CCT TAA GGT ATT ACT CAG CT .

The EcoRI site allows the insertion of heterologous sequences which, when expressed from a vector containing 30 the cassette, yield polypeptides which are fused to superoxide dismutase via an oligopeptide linker containing the amino acid sequence:

-asn-leu-gly-ile-arg-.

A sample of pS356 has been deposited on 29 April 1988 under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20853, and has been assigned Accession No. 67683. The terms and conditions for availability and access to the deposit, and for maintenance of the deposit are the same as those specified in Section II.A., for strains containing NANBV-cDNAs. This deposit is intended for convenience only, and is not required to practice the present invention in view of the description here. The deposited material is hereby incorporated herein by reference.

After recombinants containing the C100 cDNA insert in the correct orientation were isolated, the expression cassette containing the C100 cDNA was excised from pS3-56_{C100} with BamHI, and a fragment of ~3400bp which contains the cassette was isolated and purified. This fragment was then inserted into the BamHI site of the yeast vector pAB24.

Plasmid pAB24, the significant features of which are shown in Fig. 35, is a yeast shuttle vector which contains the complete 2 micron sequence for replication [Broach (1981)] and pBR322 sequences. It also contains the yeast URA3 gene derived from plasmid YEp24 [Botstein et al. (1979)], and the yeast LEU^{2d} gene derived from plasmid pC1/1. EPO Pub. No. 116,201. Plasmid pAB24 was constructed by digesting YEp24 with EcoRI and religating the vector to remove the partial 2 micron sequences. The resulting plasmid, YEP24deltaRI, was linearized by digestion with Clal and ligated with the complete 2 micron plasmid which had been linearized with Clal. The resulting plasmid, pCBou, was then digested with XbaI and the 8605 bp vector fragment was gel isolated. This isolated XbaI fragment was ligated with a 4460 bp XbaI fragment containing the LEU^{2d} gene isolated from pC1/1;

the orientation of the LEU^{2d} gene is in the same direction as the URA3 gene. Insertion of the expression was in the unique BamHI site of the pBR322 sequence, thus interrupting the gene for bacterial resistance to

5 tetracycline.

The recombinant plasmid which contained the SOD-C100 expression cassette, pAB24C100-3, was transformed into yeast strain JSC 308, as well as into other yeast strains. The cells were transformed as described by
10 Hinnen et al. (1978), and plated onto ura-selective plates. Single colonies were inoculated into leu-selective media and grown to saturation. The culture was induced to express the SOD-C100 polypeptide (called C100-3) by growth in YEP containing 1% glucose.

15 Strain JSC 308 is of the genotype MAT α , leu2, ura3(del) DM15 (GAP/ADR1) integrated at the ADR1 locus. In JSC 308, over-expression of the positive activator gene product, ADR1, results in hyperderepression (relative to an ADR1 wild type control) and significantly higher yields
20 of expressed heterologous proteins when such proteins are synthesized via an ADH2 UAS regulatory system. The construction of the yeast strain JSC 308 is disclosed in copending application, U.S. Serial No. (Attorney Docket No. 2300-0229), filed concurrently herewith, and which is
25 hereby incorporated herein by reference. A sample of JSC 308 has been deposited on 5 May 1988 with the ATCC under the conditions of the Budapest Treaty, and has been assigned Accession No. 20879. The terms and conditions for availability and access to the deposit, and for
30 maintenance of the deposit are the same as those specified in Section II.A., for strains containing HCV cDNAs.

35 The complete C100-3 fusion polypeptide encoded in pAB24C100-3 should contain 154 amino acids of human SOD at the amino-terminus, 5 amino acid residues derived from the synthetic adaptor containing the EcoRI site, 363 amino

acid residues derived from C100 cDNA, and 5 carboxy-terminal amino acids derived from the MS2 nucleotide sequence adjoining the HCV cDNA sequence in clone 32. (See Section IV.A.7.) The putative amino acid sequence of 5 the carboxy-terminus of this polypeptide, beginning at the penultimate Ala residue of SOD, is shown in Fig. 36; also shown is the nucleotide sequence encoding this portion of the polypeptide.

10 10 IV.B.5. Identification of the Polypeptide Encoded within C100 as an NANBH Associated Antigen

The C100-3 fusion polypeptide expressed from plasmid pAB24C100-3 in yeast strain JSC 308 was characterized with respect to size, and the polypeptide encoded 15 within C100 was identified as an NANBH-associated antigen by its immunological reactivity with serum from a human with chronic NANBH.

The C100-3 polypeptide, which was expressed as described in Section IV.B.4., was analyzed as follows.

20 Yeast JSC 308 cells were transformed with pAB24, or with pAB24C100-3, and were induced to express the heterologous plasmid encoded polypeptide. The induced yeast cells in 1 ml of culture ($OD_{650\text{ nm}} \sim 20$) were pelleted by centrifugation at 10,000 rpm for 1 minute, and were lysed 25 by vortexing them vigorously ($10 \times 1\text{ min}$) with 2 volumes of solution and 1 volume of glass beads (0.2 millimicron diameter). The solution contained 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1mM phenylmethylsulphonyl fluoride (PMSF), and 1 microgram/ml pepstatin. Insoluble material in the 30 lysate, which includes the C100-3 polypeptide, was collected by centrifugation (10,000 rpm for 5 minutes), and was dissolved by boiling for 5 minutes in Laemmli SDS sample buffer. [See Laemmli (1970)]. An amount of 35 polypeptides equivalent to that in 0.3 ml of the induced yeast culture was subjected to electrophoresis through 10%

polyacrylamide gels in the presence of SDS according to Laemmli (1970). Protein standards were co-electrophoresed on the gels. Gels containing the expressed polypeptides were either stained with Coomassie brilliant blue, or were 5 subjected to "Western" blotting as described in Section IV.B.2., using serum from a patient with chronic NANBH to determine the immunological reactivity of the polypeptides expressed from pAB24 and from pAB24C100-3.

The results are shown in Fig. 37. In Fig. 37A 10 the polypeptides were stained with Coomassie brilliant blue. The insoluble polypeptide(s) from JSC 308 transformed with pAB24 and from two different colonies of JSC transformed with pAB24C100-3 are shown in lane 1 (pAB24), and lanes 2 and 3, respectively. A comparison of lanes 2 15 and 3 with lane 1 shows the induced expression of a polypeptide corresponding to a molecular weight of ~54,000 daltons from JSC 308 transformed with pAB24C100-3, which is not induced in JSC 308 transformed with pAB24. This polypeptide is indicated by the arrow.

Fig. 37B shows the results of the Western blots 20 of the insoluble polypeptides expressed in JSC 308 transformed with pAB24 (lane 1), or with pAB24C100-3 (lane 2). The polypeptides expressed from pAB24 were not immunologically reactive with serum from a human with NANBH. 25 However, as indicated by the arrow, JSC 308 transformed with pAB24C100-3 expressed a polypeptide of ~54,000 dalton molecular weight which did react immunologically with the human NANBH serum. The other immunologically reactive polypeptides in lane 2 may be degradation and/or aggregation products of this ~54,000 dalton polypeptide. 30

IV.B.6. Purification of Fusion Polypeptide C100-3

The fusion polypeptide, C100-3, comprised of SOD at the N-terminus and in-frame C100 HCV-polypeptide at the 35 C-terminus was purified by differential extraction of the

insoluble fraction of the extracted host yeast cells in which the polypeptide was expressed.

The fusion polypeptide, C100-3, was expressed in yeast strain JSC 308 transformed with pAB24C100-3, as described in Section IV.B.4. The yeast cells were then lysed by homogenization, the insoluble material in the lysate was extracted at pH 12.0, and C100-3 in the remaining insoluble fraction was solubilized in buffer containing SDS.

10 The yeast lysate was prepared essentially according to Nagahuma et al. (1984). A yeast cell suspension was prepared which was 33% cells (v/v) suspended in a solution (Buffer A) containing 20 mM Tris HCl, pH 8.0, 1 mM dithiothreitol, and 1 mM phenylmethylsulfonylfluoride (PMSF). An aliquot of the suspension (15 ml) was mixed with an equal volume of glass beads (0.45-0.50 mm diameter), and the mixture was vortexed at top speed on a Super Mixer (Lab Line Instruments, Inc.) for 8 min. The homogenate and glass beads were separated, and the glass 20 beads were washed 3 times with the same volume of Buffer A as the original packed cells. After combining the washes and homogenate, the insoluble material in the lysate was obtained by centrifuging the homogenate at 7,000 x g for 15 minutes at 4°C, resuspending the pellets in Buffer A 25 equal to twice the volume of original packed cells, and re-pelleting the material by centrifugation at 7,000 x g for 15 min. This washing procedure was repeated 3 times.

30 The insoluble material from the lysate was extracted at pH 12.0 as follows. The pellet was suspended in buffer containing 0.5 M NaCl, 1 mM EDTA, where the suspending volume was equal to 1.8 times the of the original packed cells. The pH of the suspension was adjusted by adding 0.2 volumes of 0.4 M Na phosphate buffer, pH 12.0. After mixing, the suspension was 35 centrifuged at 7,000 x g for 15 min at 4°C, and the super-

natant removed. The extraction was repeated 2 times. The extracted pellets were washed by suspending them in 0.5 M NaCl, 1 mM EDTA, using a suspension volume equal to two volumes of the original packed cells, followed by

5 centrifugation at 7,000 x g for 15 min at 4°C.

- The C100-3 polypeptide in the extracted pellet was solubilized by treatment with SDS. The pellets were suspended in Buffer A equal to 0.9 volumes of the original packed cell volume, and 0.1 volumes of 2% SDS was added.
- 10 After the suspension was mixed, it was centrifuged at 7,000 x g for 15 min at 4°C. The resulting pellet was extracted 3 more times with SDS. The resulting supernatants, which contained C100-3 were pooled.

This procedure purifies C100-3 more than 10-fold from the insoluble fraction of the yeast homogenate, and the recovery of the polypeptide is greater than 50%.

The purified preparation of fusion polypeptide was analyzed by polyacrylamide gel electrophoresis according to Laemmli (1970). Based upon this analysis, the 20 polypeptide was greater than 80% pure, and had an apparent molecular weight of ~54,000 daltons.

IV.C. Identification of RNA in Infected Individuals Which Hybridizes to HCV cDNA.

- 25 IV.C.1. Identification of RNA in the Liver of a Chimpanzee With NANBH Which Hybridizes to HCV cDNA.
- RNA from the liver of a chimpanzee which had NANBH was shown to contain a species of RNA which hybridized to the HCV cDNA contained within clone 81 by Northern blotting, as follows.
- RNA was isolated from a liver biopsy of the chimpanzee from which the high titer plasma was derived (see Section IV.A.1.) using techniques described in
- 35 Maniatis et al. (1982) for the isolation of total RNA from

mammalian cells, and for its separation into poly A⁺ and poly A⁻ fractions. These RNA fractions were subjected to electrophoresis on a formaldehyde/agarose gel (1% w/v), and transferred to nitrocellulose. (Maniatis et al. 5 (1982)). The nitrocellulose filters were hybridized with radiolabeled HCV cDNA from clone 81 (see Fig. 4 for the nucleotide sequence of the insert.) To prepare the radiolabeled probe, the HCV cDNA insert isolated from clone 81 was radiolabeled with ³²P by nick translation 10 using DNA Polymerase I (Maniatis et al. (1982)). Hybridization was for 18 hours at 42°C in a solution containing 10% (w/v) Dextran sulphate, 50% (w/v) deionized formamide, 750 mM NaCl, 75 mM Na citrate, 20 mM Na₂HPO₄, pH 6.5, 0.1% SDS, 0.02% (w/v) bovine serum albumin (BSA), 15 0.02% (w/v) Ficoll-400, 0.02% (w/v) polyvinylpyrrolidone, 100 micrograms/ml salmon sperm DNA which had been sheared by sonication and denatured, and 10⁶ CPM/ml of the nick-translated cDNA probe.

An autoradiograph of the probed filter is shown 20 in Fig. 38. Lane 1 contains ³²P-labeled restriction fragment markers. Lanes 2-4 contain chimpanzee liver RNA as follows: lane 2 contains 30 micrograms of total RNA; lane 3 contains 30 micrograms of poly A- RNA; and lane 4 contains 20 micrograms of poly A+ RNA. As shown in Fig. 25 38, the liver of the chimpanzee with NANBH contains a heterogeneous population of related poly A+ RNA molecules which hybridizes to the HCV cDNA probe, and which appears to be from about 5000 nucleotides to about 11,000 nucleotides in size. This RNA, which hybridizes to the 30 HCV cDNA, could represent viral genomes and/or specific transcripts of the viral genome.

The experiment described in Section IV.C.2., infra, is consistent with the suggestion that HCV contains an RNA genome.

IV.C.2. Identification of HCV Derived RNA in Serum from Infected Individuals.

Nucleic acids were extracted from particles isolated from high titer chimpanzee NANBH plasma as described in Section IV.A.1.. Aliquots (equivalent to 1 ml of original plasma) of the isolated nucleic acids were resuspended in 20 microliters 50 mM Hepes, pH 7.5, 1 mM EDTA and 16 micrograms/ml yeast soluble RNA. The samples were denatured by boiling for 5 minutes followed by immediate freezing, and were treated with RNase A (5 microliters containing 0.1 mg/ml RNase A in 25 mM EDTA, 40 mM Hepes, pH 7.5) or with DNase I (5 microliters containing 1 unit DNase I in 10 mM MgCl₂, 25 mM Hepes, pH 7.5); control samples were incubated without enzyme. Following incubation, 230 microliters of ice-cold 2XSSC containing 2 micrograms/ml yeast soluble RNA was added, and the samples were filtered on a nitrocellulose filter. The filters were hybridized with a cDNA probe from clone 81, which had been ³²P-labeled by nick-translation. Fig. 39 shows an autoradiograph of the filter. Hybridization signals were detected in the DNase treated and control samples (lanes 2 and 1, respectively), but were not detected in the RNase treated sample (lane 3). Thus, since RNase A treatment destroyed the nucleic acids isolated from the particles, and DNase I treatment had no effect, the evidence strongly suggests that the HCV genome is composed of RNA.

IV.C.3. Detection of Amplified HCV Nucleic Acid Sequences derived from HCV Nucleic Acid Sequences in Liver and plasma Specimens from Chimpanzees with NANBH

HCV nucleic acids present in liver and plasma of chimpanzees with NANBH, and in control chimpanzees, were amplified using essentially the polymerase chain reaction (PCR) technique described by Saiki et al. (1986). The primer oligonucleotides were derived from the HCV cDNA

sequences in clone 81, or clones 36 and 37. The amplified sequences were detected by gel electrophoresis and Southern blotting, using as probes the appropriate cDNA oligomer with a sequence from the region between, but not 5 including, the two primers.

Samples of RNA containing HCV sequences to be examined by the amplification system were isolated from liver biopsies of three chimpanzees with NANBH, and from two control chimpanzees. The isolation of the RNA fraction was by the guanidinium thiocyanate procedure 10 described in Section IV.C.1.

Samples of RNA which were to be examined by the amplification system were also isolated from the plasmas of two chimpanzees with NANBH, and from one control 15 chimpanzee, as well as from a pool of plasmas from control chimpanzees. One infected chimpanzee had a CID/ml equal to or greater than 10^6 , and the other infected chimpanzee had a CID/ml equal to or greater than 10^5 .

The nucleic acids were extracted from the plasma 20 as follows. Either 0.1 ml or 0.01 ml of plasma was diluted to a final volume of 1.0 ml, with a TENB/proteinase K/SDS solution (0.05 M Tris-HCL, pH 8.0, 0.001 M EDTA, 0.1 M NaCl, 1 mg/ml Proteinase K, and 0.5% SDS) containing 10 micrograms/ml polyadenylic acid, and 25 incubated at 37°C for 60 minutes. After this proteinase K digestion, the resultant plasma fractions were deproteinized by extraction with TE (10.0 mM Tris-HCl, pH 8.0, 1 mM EDTA) saturated phenol. The phenol phase was separated by centrifugation, and was reextracted with TENB 30 containing 0.1% SDS. The resulting aqueous phases from each extraction were pooled, and extracted twice with an equal volume of phenol/chloroform/isoamyl alcohol [1:1(99:2)], and then twice with an equal volume of a 99:1 mixture of chloroform/isoamyl alcohol. Following phase 35 separation by centrifugation, the aqueous phase was

brought to a final concentration of 0.2 M Na Acetate, and the nucleic acids were precipitated by the addition of two volumes of ethanol. The precipitated nucleic acids were recovered by ultracentrifugation in a SW 41 rotor at 38 K,
5 for 60 minutes at 4°C.

- In addition to the above, the high titer chimpanzee plasma and the pooled control plasma alternatively were extracted with 50 micrograms of poly A carrier by the procedure of Chomczynski and Sacchi (1987).
10 This procedure uses an acid guanidinium thiocyanate extraction. RNA was recovered by centrifugation at 10,000 RPM for 10 minutes at 4°C in an Eppendorf microfuge.

On two occasions, prior to the synthesis of cDNA in the PCR reaction, the nucleic acids extracted from
15 plasma by the proteinase K/SDS/phenol method were further purified by binding to and elution from S and S Elutip-R Columns. The procedure followed was according to the manufacturer's directions.

The cDNA used as a template for the PCR reaction
20 was derived from the nucleic acids (either total nucleic acids or RNA) prepared as described above. Following ethanol precipitation, the precipitated nucleic acids were dried, and resuspended in DEPC treated distilled water. Secondary structures in the nucleic acids were disrupted
25 by heating at 65°C for 10 minutes, and the samples were immediately cooled on ice. cDNA was synthesized using 1 to 3 micrograms of total chimpanzee RNA from liver, or from nucleic acids (or RNA) extracted from 10 to 100 microliters of plasma. The synthesis utilized reverse transcriptase, and was in a 25 microliter reaction, using
30 the protocol specified by the manufacturer, BRL. The primers for cDNA synthesis were those also utilized in the PCR reaction, described below. All reaction mixtures for cDNA synthesis contained 23 units of the RNAase inhibitor,
35 RNasin® (Fisher/Promega). Following cDNA synthesis, the

reaction mixtures were diluted with water, boiled for 10 minutes, and quickly chilled on ice.

The PCR reactions were performed essentially according to the manufacturer's directions (Cetus-Perkin-
5 Elmer), except for the addition of 1 microgram of RNase A. The reactions were carried out in a final volume of 100 microliters. The PCR was performed for 35 cycles, utilizing a regimen of 37°C, 72°C, and 94°C.

The primers for cDNA synthesis and for the PCR
10 reactions were derived from the HCV cDNA sequences in either clone 81, clone 36, or clone 37b. (The HCV cDNA sequences of clones 81, 36, and 37b are shown in Figs. 4, 5, and 10, respectively.) The sequences of the two 16-mer primers derived from clone 81 were:

15

5' CAA TCA TAC CTG ACA G 3'

and

5' GAT AAC CTC TGC CTG A 3'.

20 The sequence of the primer from clone 36 was:

5' GCA TGT CAT GAT GTA T 3'.

The sequence of the primer from clone 37b was:

25

5' ACA ATA CGT GTG TCA C 3'.

In the PCR reactions, the primer pairs consisted of either the two 16-mers derived from clone 81, or the 16-mer from
30 clone 36 and the 16-mer from clone 37b.

The PCR reaction products were analyzed by separation of the products by alkaline gel electrophoresis, followed by Southern blotting, and detection of the amplified HCV-cDNA sequences with a ³²P-labeled internal oligonucleotide probe derived from a
35

region of the HCV cDNA which does not overlap the primers. The PCR reaction mixtures were extracted with phenol/chloroform, and the nucleic acids precipitated from the aqueous phase with salt and ethanol. The precipitated 5 nucleic acids were collected by centrifugation, and dissolved in distilled water. Aliquots of the samples were subjected to electrophoresis on 1.8% alkaline agarose gels. Single stranded DNA of 60, 108, and 161 nucleotide lengths were co-electrophoresed on the gels as molecular 10 weight markers. After electrophoresis, the DNAs in the gel were transferred onto Biorad Zeta Probe⁺ paper. Prehybridization and hybridization, and wash conditions were those specified by the manufacturer (Biorad).

The probes used for the hybridization-detection 15 of amplified HCV cDNA sequences were the following. When the pair of PCR primers were derived from clone 81, the probe was an 108-mer with a sequence corresponding to that which is located in the region between the sequences of the two primers. When the pair of PCR primers were 20 derived from clones 36 and 37b, the probe was the nick-translated HCV cDNA insert derived from clone 35. The primers are derived from nucleotides 155-170 of the clone 37b insert, and 206-268 of the clone 36 insert. The 3'-end of the HCV cDNA insert in clone 35 overlaps 25 nucleotides 1-186 of the insert in clone 36; and the 5'-end of clone 35 insert overlaps nucleotides 207-269 of the insert in clone 37b. (Compare Figs. 5, 8 and 10.) Thus, the cDNA insert in clone 35 spans part of the region between the sequences of the clone 36 and 37b derived 30 primers, and is useful as a probe for the amplified sequences which include these primers.

Analysis of the RNA from the liver specimens was according to the above procedure utilizing both sets of primers and probes. The RNA from the liver of the three 35 chimpanzees with NANBH yielded positive hybridization

results for amplification sequences of the expected size (161 and 586 nucleotides for 81 and 36 and 37b, respectively), while the control chimpanzees yielded negative hybridization results. The same results were

5 achieved when the experiment was repeated three times.

Analysis of the nucleic acids and RNA from plasma was also according to the above procedure utilizing the primers and probe from clone 81. The plasmas were from two chimpanzees with NANBH, from a control

10 chimpanzee, and pooled plasmas from control chimpanzees. Both of the NANBH plasmas contained nucleic acids/RNA which yielded positive results in the PCR amplified assay, while both of the control plasmas yielded negative results. These results have been repeatably obtained

15 several times.

IV.D. Radioimmunoassay for Detecting HCV Antibodies in Serum from Infected Individuals

Solid phase radioimmunoassays to detect antibodies to HCV antigens were developed based upon Tsu and Herzenberg (1980). Microtiter plates (Immulon 2, Removewell strips) are coated with purified polypeptides containing HCV epitopes. The coated plates are incubated with either human serum samples suspected of containing

20 antibodies to the HCV epitopes, or to appropriate controls. During incubation, antibody, if present, is immunologically bound to the solid phase antigen. After removal of the unbound material and washing of the microtiter plates, complexes of human antibody-NANBV

25 antigen are detected by incubation with ¹²⁵I-labeled sheep anti-human immunoglobulin. Unbound labeled antibody is removed by aspiration, and the plates are washed. The radioactivity in individual wells is determined; the amount of bound human anti-HCV antibody is proportional to

30 the radioactivity in the well.

35

IV.D.1. Purification of Fusion Polypeptide SOD-NANB₅₋₁₋₁

The fusion polypeptide SOD-NANB₅₋₁₋₁, expressed in recombinant bacteria as described in Section IV.B.1., was purified from the recombinant E. coli by differential extraction of the cell extracts with urea, followed by chromatography on anion and cation exchange columns as follows.

Thawed cells from 1 liter of culture were resuspended in 10 ml of 20% (w/v) sucrose containing 0.01M Tris HCl, pH 8.0, and 0.4 ml of 0.5M EDTA, pH 8.0 was added. After 5 minutes at 0°C, the mixture was centrifuged at 4,000 x g for 10 minutes. The resulting pellet was suspended in 10 ml of 25% (w/v) sucrose containing 0.05 M Tris HCl, pH 8.0, 1 mM phenylmethylsulfonylfluoride (PMSF) and 1 microgram/ml pepstatin A, followed by addition of 0.5 ml lysozyme (10 mg/ml) and incubation at 0°C for 10 minutes. After the addition of 10 ml 1% (v/v) Triton X-100 in 0.05 M Tris HCl, pH 8.0, 1 mM EDTA, the mixture was incubated an additional 10 min at 0°C with occasional shaking. The resulting viscous solution was homogenized by passage 6 times through a sterile 20-gauge hypodermic needle, and centrifuged at 13,000 x g for 25 minutes. The pelleted material was suspended in 5 ml of 0.01 M Tris HCl pH 8.0, and the suspension centrifuged at 4,000 x g for 10 minutes. The pellet, which contained SOD-NANB₅₋₁₋₁ fusion protein, was dissolved in 5 ml of 6 M urea in 0.02 M Tris HCl, pH 8.0, 1 mM dithiothreitol (Buffer A), and was applied to a column of Q-Sepharose Fast Flow equilibrated with Buffer A. Polypeptides were eluted with a linear gradient of 0.0 to 0.3 M NaCl in Buffer A. After elution, fractions were analyzed by polyacrylamide gel electrophoresis in the presence of SDS to determine their content of SOD-NANB₅₋₁₋₁. Fractions containing this polypeptide were pooled, and dialyzed against 6 M urea in

0.02 M sodium phosphate buffer, pH 6.0, 1 mM dithiothreitol (Buffer B). The dialyzed sample was applied on a column of S-Sepharose Fast Flow equilibrated with Buffer B, and polypeptides eluted with a linear 5 gradient of 0.0 to 0.3 M NaCl in Buffer B. The fractions were analyzed by polyacrylamide gel electrophoresis for the presence of SOD-NANB₅₋₁₋₁, and the appropriate fractions were pooled.

The final preparation of SOD-NANB₅₋₁₋₁ 10 polypeptide was examined by electrophoresis on polyacrylamide gels in the presence of SDS. Based upon this analysis, the preparation was more than 80% pure.

IV.D.2. Purification of Fusion Polypeptide SOD-NANB₈₁.

15 The fusion polypeptide SOD-NANB₈₁, expressed in recombinant bacteria as described in Section IV.B.2., was purified from recombinant E. coli by differential extraction of the cell extracts with urea, followed by chromatography on anion and cation exchange columns 20 utilizing the procedure described for the isolation of fusion polypeptide SOD-NANB₅₋₁₋₁ (See Section IV.D.1.).

The final preparation of SOD-NANB₈₁ polypeptide was examined by electrophoresis on polyacrylamide gels in the presence of SDS. Based upon this analysis, the 25 preparation was more than 50% pure.

IV.D.3. Detection of Antibodies to HCV Epitopes by Solid Phase Radioimmunoassay.

Serum samples from 32 patients who were 30 diagnosed as having NANBH were analyzed by radioimmunoassay (RIA) to determine whether antibodies to HCV epitopes present in fusion polypeptides SOD-NANB₅₋₁₋₁ and SOD-NANB₈₁ were detected.

Microtiter plates were coated with SOD-NANB₅₋₁₋₁ 35 or SOD-NANB₈₁, which had been partially purified according

to Sections IV.D.1. and IV.D.2., respectively. The assays were conducted as follows.

- One hundred microliter aliquots containing 0.1 to 0.5 micrograms of SOD-NANB₅₋₁₋₁ or SOD-NANB₈₁ in 0.125 M Na borate buffer, pH 8.3, 0.075 M NaCl (BBS) was added to each well of a microtiter plate (Dynatech Immulon 2 Removawell Strips). The plate was incubated at 4°C overnight in a humid chamber, after which, the protein solution was removed and the wells washed 3 times with BBS containing 0.02% Triton X-100 (BBST). To prevent non-specific binding, the wells were coated with bovine serum albumin (BSA) by addition of 100 microliters of a 5 mg/ml solution of BSA in BBS followed by incubation at room temperature for 1 hour; after this incubation the BSA solution was removed. The polypeptides in the coated wells were reacted with serum by adding 100 microliters of serum samples diluted 1:100 in 0.01M Na phosphate buffer, pH 7.2, 0.15 M NaCl (PBS) containing 10 mg/ml BSA, and incubating the serum containing wells for 1 hr at 37°C.
- After incubation, the serum samples were removed by aspiration, and the wells were washed 5 times with BBST. Anti-NANB₅₋₁₋₁ and Anti-NANB₈₁ bound to the fusion polypeptides was determined by the binding of ¹²⁵I-labeled F'(ab)₂ sheep anti-human IgG to the coated wells.
- Aliquots of 100 microliters of the labeled probe (specific activity 5-20 microcuries/microgram) were added to each well, and the plate's were incubated at 37°C for 1 hour, followed by removal of excess probe by aspiration, and 5 washes with BBST. The amount of radioactivity bound in each well was determined by counting in a counter which detects gamma radiation.

The results of the detection of anti-NANB₅₋₁₋₁ and anti-NANB₈₁ in individuals with NANBH is presented in Table 1.

Table 1
Detection of Anti-S-1-1 and Anti-81 in Sera of
NANB, HAV and HBV Hepatitis Patients

	Patient Reference Number	Diagnosis	S/N	Anti-S-1-1	Anti-81
5	1. 28 ¹	Chronic NANB, IVD ²		0.77	4.20
		Chronic NANB, IVD		1.14	5.14
		Chronic NANB, IVD		2.11	4.05
10	2. 29 ¹	AVH ³ , NANB, Sporadic		1.09	1.05
		Chronic, NANB		33.89	11.39
		Chronic, NANB		36.22	13.67
15	3. 30 ¹	AVH, NANB, IVD		1.90	1.54
		Chronic NANB, IVD		34.17	30.28
		Chronic NANB, IVD		32.45	30.84
20	4. 31	Chronic NANB, PT ⁴		16.09	8.05
	5. 32 ¹	Late AVH NANB, IVD		0.69	0.94
25	6. 33 ¹	AVH, NANB, IVD		1.66	1.96
		AVH, NANB, IVD		1.53	0.56
	7. 34 ¹	Chronic NANB, PT		34.40	7.55
30		Chronic NANB, PT		45.55	13.11
		Chronic NANB, PT		41.58	13.45
		Chronic NANB, PT		44.20	15.48
	8. 35 ¹	AVH NANB, IVD "Healed" recent NANB, AVH		31.92 6.87	31.95 4.45
35	9. 36	Late AVH NANB PT		11.84	5.79
	10. 37	AVH NANB, IVD		6.52	1.33
40	11. 38	Late AVH NANB, PT		39.44	39.18
	12. 39	Chronic NANB, PT		42.22	37.54
	13. 40	AVH, NANB, PT		1.35	1.17
45	14. 41	Chronic NANB? PT		0.35	0.28

Patient Reference Number	Diagnosis	S/N	
		Anti-S-1-1	Anti-81
	15. 42 AVH, NANB, IVD	6.25	2.34
5	16. 43 Chronic NANB, PT	0.74	0.61
	17. 44 AVH, NANB, PT	5.40	1.83
	18. 45 Chronic, NANB, PT	0.52	0.32
	19. 46 AVH, NANB	23.35	4.45
10	20. 47 AVH, Type A	1.60	1.35
	21. 48 AVH, Type A	1.30	0.66
	22. 49 AVH, Type A	1.44	0.74
	23. 50 Resolved Recent AVH, Type A	0.48	0.56
15	24. 51 AVH, Type A Resolved AVH, Type A	0.68 0.80	0.64 0.65
	25. 52 Resolved Recent AVH, Type A Resolved Recent AVH, Type A	1.38 0.80	1.04 0.65
20	26. 53 AVH, Type A Resolved Recent AVH, Type A	1.85 1.02	1.16 0.88
	27. 54 AVH, Type A	1.35	0.74
	28. 55 Late AVH, HBV	0.58	0.55
25	29. 56 Chronic HBV	0.84	1.06
	30. 57 Late AVH, HBV	3.20	1.60
	31. 58 Chronic HBV	0.47	0.46
	32. 59 ¹ AVH, HBV Healed AVH, HBV	0.73 0.43	0.60 0.44
30	33. 60 ¹ AVH, HBV Healed AVH, HBV	1.06 0.75	0.92 0.68

5	Patient Reference Number	Diagnosis	S/N	
			Anti-5-1-1	Anti-81
	34. 61 ¹	AVH, HBV Healed AVH, HBV	1.66 0.63	0.61 0.36
	35. 62 ¹	AVH, HBV Healed AVH, HBV	1.02 0.41	0.73 0.42
10	36. 63 ¹	AVH, HBV Healed AVH, HBV	1.24 1.55	1.31 0.45
	37. 64 ¹	AVH, HBV Healed AVH, HBV	0.82 0.53	0.79 0.37
	38. 65 ¹	AVH, HBV Healed AVH, HBV	0.95 0.70	0.92 0.50
15	39. 66 ¹	AVH, HBV Healed AVH, HBV	1.03 1.71	0.68 1.39

¹ Sequential serum samples available from these patients

² IVUD=Intravenous Drug User

³ AVH=Acute viral hepatitis

⁴ PT=Post transfusion

As seen in Table 1, 19 of 32 sera from patients diagnosed as having NANBH were positive with respect to antibodies directed against HCV epitopes present in SOD-NANB₅₋₁₋₁ and SOD-NANB₈₁.

5 However, the serum samples which were positive were not equally immunologically reactive with SOD-NANB₅₋₁₋₁ and SOD-NANB₈₁. Serum samples from patient No. 1 were positive to SOD-NANB₈₁ but not to SOD-NANB₅₋₁₋₁. Serum samples from patients number 10, 15, and 17 were positive
10 to SOD-NANB₅₋₁₋₁ but not to SOD-NANB₈₁. Serum samples from patients No. 3, 8, 11, and 12 reacted equally with both fusion polypeptides, whereas serum samples from patients No. 2, 4, 7, and 9 were 2-3 fold higher in the reaction to SOD-NANB₅₋₁₋₁ than to SOD-NANB₈₁. These
15 results suggest that NANB₅₋₁₋₁ and NANB₈₁ may contain at least 3 different epitopes; i.e., it is possible that each polypeptide contains at least 1 unique epitope, and that the two polypeptides share at least 1 epitope.

20 . IV.D.4. Specificity of the Solid Phase RIA for NANBH

The specificity of the solid phase RIAs for NANBH was tested by using the assay on serum from patients infected with HAV or with HBV and on sera from control individuals. The assays utilizing partially purified SOD-
25 NANB₅₋₁₋₁ and SOD-NANB₈₁ were conducted essentially as described in Section IV.D.3, except that the sera was from patients previously diagnosed as having HAV or HBV, or from individuals who were blood bank donors. The results for sera from HAV and HBV infected patients are presented
30 in table 1. The RIA was tested using 11 serum specimens from HAV infected patients, and 20 serum specimens from HBV infected patients. As shown in table 1, none of these sera yielded a positive immunological reaction with the fusion polypeptides containing BB-NANBV epitopes.

The RIA using the NANB₅₋₁₋₁ antigen was used to determine immunological reactivity of serum from control individuals. Out of 230 serum samples obtained from the normal blood donor population, only 2 yielded positive reactions in the RIA (data not shown). It is possible that the two blood donors from whom these serum samples originated had previously been exposed to HCV.

10 IV.D.5. Reactivity of NANB₅₋₁₋₁ During the Course of NANBH Infection.

15 The presence of anti-NANB₅₋₁₋₁ antibodies during the course of NANBH infection of 2 patients and 4 chimpanzees was followed using RIA as described in Section
IV.D.3. In addition the RIA was used to determine the presence or absence of anti-NANB₅₋₁₋₁ antibodies during the course of infection of HAV and HBV in infected chimpanzees.

20 The results, which are presented in Table 2, show that with chimpanzees and with humans, anti-NANB₅₋₁₋₁ antibodies were detected following the onset of the acute phase of NANBH infection. Anti-NANB₅₋₁₋₁ antibodies were not detected in serum samples from chimpanzees infected with either HAV or HBV. Thus anti-NANB₅₋₁₋₁ antibodies serve as a marker for an individual's exposure to HCV.

Table 2
Seroconversion in Sequential Serum Samples from
Hepatitis Patients and Chimpanzees Using S-I-I Antigen

	PATIENT/ Chimp	Sample Date (Days) (0=inoculation day)	Hepatitis Viruses	Anti-S-I-I (S/N)	ALT (mu/ml)
5	Patient 29	T6	NANB	1.09	1180
		T+180		33.89	425
		T+208		36.22	--
10	Patient 30	T	NANB	1.90	1830
		T+307		34.17	290
		T+799		32.45	276
15	Chimp 1	0	NANB	0.87	9
		76		0.93	71
		118		23.67	19
		154		32.41	--
	Chimp 2	0	NANB	1.00	5
20		21		1.08	52
		73		4.64	13
		138		25.01	--
Chimp 3	0	NANB	1.08	8	
	43		1.44	205	
	53		1.82	14	
	159		11.87	6	
25	Chimp 4	-3	NANB	1.12	11
		55		1.25	132
		83		6.60	--
		140		17.51	--
	Chimp 5	0	HAV	1.50	4
30		25		2.39	147
		40		1.92	18
		268		1.53	5
Chimp 6	-8	HAV	0.85	--	
	15		--	106	
	41		0.81	10	
	129		1.33	--	

30

35

- 119 -

Patient/ Chimp	Sample Date (0=inoculation day)	Hepatitis Viruses	Anti-5-l-1 (S/N)	ALT (mu/ml)
Chimp 7	0	HAV	1.17	7
	22		1.60	83
	115		1.55	5
	139		1.60	--
Chimp 8	0	HAV	0.77	15
	26		0.98	130
	74		1.77	8
	205		1.27	5
Chimp 9	-290	HBV	1.74	--
	379		3.29	9
	435		2.77	6
Chimp 10	0	HBV	2.35	8
	111-118 (pool)		2.74	96-156 (pool)
	205		2.05	9
	240		1.78	13
Chimp 11	0	HBV	1.82	11
	28-56 (pool)		1.26	8-100 (pool)
	169		--	9
	223		0.52	10

*T=day of initial sampling

IV.E. Purification of Polyclonal Serum Antibodies to
NANB₅₋₁₋₁

On the basis of the specific immunological reactivity of the SOD-NANB₅₋₁₋₁ polypeptide with the anti-bodies in serum samples from patients with NANBH, a method was developed to purify serum antibodies which react immunologically with the epitope(s) in NANB₅₋₁₋₁. This method utilizes affinity chromatography. Purified SOD-NANB₅₋₁₋₁ polypeptide (see Section IV.D.1) was attached to an insoluble support; the attachment is such that the immobilized polypeptide retains its affinity for antibody to NANB₅₋₁₋₁. Antibody in serum samples is absorbed to the matrix-bound polypeptide. After washing to remove non-specifically bound materials and unbound materials, the bound antibody is released from the bound SOD-HCV polypeptide by change in pH, and/or by chaotropic reagents, for example, urea.

Nitrocellulose membranes containing bound SOD-NANB₅₋₁₋₁ were prepared as follows. A nitrocellulose membrane, 2.1 cm Sartorius of 0.2 micron pore size, was washed for 3 minutes three times with BBS. SOD-NANB₅₋₁₋₁ was bound to the membrane by incubation of the purified preparation in BBS at room temperature for 2 hours; alternatively it was incubated at 4°C overnight. The solution containing unbound antigen was removed, and the filter was washed three times with BBS for three minutes per wash. The remaining active sites on the membrane were blocked with BSA by incubation with a 5 mg/ml BSA solution for 30 minutes. Excess BSA was removed by washing the membrane with 5 times with BBS and 3 times with distilled water. The membrane containing the viral antigen and BSA was then treated with 0.05 M glycine hydrochloride, pH 2.5, 0.10 M NaCl (GlyHCl) for 15 minutes, followed by 3 three minute washes with PBS.

- Polyclonal anti-NANB₅₋₁₋₁ antibodies were isolated by incubating the membranes containing the fusion polypeptide with serum from an individual with NANBH for 2 hours. After the incubation, the filters were washed 5 times with BBS, and twice with distilled water. Bound antibodies were then eluted from each filter with 5 elutions of GlyHCl, at 3 minutes per elution. The pH of the eluates was adjusted to pH 8.0 by collecting each eluate in a test tube containing 2.0 M Tris HCl, pH 8.0.
- 10 Recovery of the anti-NANB₅₋₁₋₁ antibody after affinity chromatography is approximately 50%.

The nitrocellulose membranes containing the bound viral antigen can be used several times without appreciable decrease in binding capacity. To reuse the membranes, after the antibodies have been eluted the membranes are washed with BBS three times for 3 minutes. They are then stored in BBS at 4°C.

- IV.F. The Capture of HCV Particles from Infected Plasma
- 20 Using Purified Human Polyclonal Anti-HCV Antibodies; Hybridization of the Nucleic Acid in the Captured Particles to HCV cDNA

- IV.F.1. The Capture of HCV Particles from Infected Plasma
- 25 Using Human Polyclonal Anti-HCV Antibodies

Protein-nucleic acid complexes present in infectious plasma of a chimpanzee with NANBH were isolated using purified human polyclonal anti-HCV antibodies which were bound to polystyrene beads.

- 30 Polyclonal anti-NANB₅₋₁₋₁ antibodies were purified from serum from a human with NANBH using the SOD-HCV polypeptide encoded in clone 5-1-1. The method for purification was that described in Section IV.E.

The purified anti-NANB₅₋₁₋₁ antibodies were bound to polystyrene beads (1/4" diameter, specular fin-

ish, Precision Plastic Ball Co., Chicago, Illinois) by incubating each at room temperature overnight with 1 ml of antibodies (1 microgram/ml in borate buffered saline, pH 8.5). Following the overnight incubation, the beads were washed once with TBST [50 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.05% (v/v) Tween 20], and then with phosphate buffered saline (PBS) containing 10 mg/ml BSA.

Control beads were prepared in an identical fashion, except that the purified anti-NANB₅₋₁₋₁ antibodies were replaced with total human immunoglobulin.

Capture of HCV from NANBH infected chimpanzee plasma using the anti-NANB₅₋₁₋₁ antibodies bound to beads was accomplished as follows. The plasma from a chimpanzee with NANBH used is described in Section IV.A.1.. An aliquot (1 ml) of the NANBV infected chimpanzee plasma was incubated for 3 hours at 37°C with each of 5 beads coated with either anti-NANB₅₋₁₋₁ antibodies, or with control immunoglobulins. The beads were washed 3 times with TBST.

IV.F.2. Hybridization of the Nucleic Acid in the Captured Particles to NANBV-cDNA

The nucleic acid component released from the particles captured with anti-NANB₅₋₁₋₁ antibodies was analyzed for hybridization to HCV cDNA derived from clone 81.

HCV particles were captured from NANBH infected chimpanzee plasma, as described in IV.F.1. To release the nucleic acids from the particles, the washed beads were incubated for 60 min. at 37°C with 0.2 ml per bead of a solution containing proteinase k (1 mg/ml), 10 mM Tris HCl, pH 7.5, 10 mM EDTA, 0.25% (w/v) SDS, 10 micrograms/ml soluble yeast RNA, and the supernatant solution was removed. The supernatant was extracted with phenol and chloroform, and the nucleic acids precipitated with ethanol overnight at -20°C. The nucleic acid precipitate

was collected by centrifugation, dried, and dissolved in 50 mM Hepes, pH 7.5. Duplicate aliquots of the soluble nucleic acids from the samples obtained from beads coated with anti-NANB₅₋₁₋₁ antibodies and with control beads 5 containing total human immunoglobulin were filtered onto to nitrocellulose filters. The filters were hybridized with a ³²P-labeled, nick-translated probe made from the purified HCV cDNA fragment in clone 81. The methods for preparing the probe and for the hybridization are 10 described in Section IV.C.1..

Autoradiographs of a probed filter containing the nucleic acids from particles captured by beads containing anti-NANB₅₋₁₋₁ antibodies are shown in Fig. 40. The extract obtained using the anti-NANB₅₋₁₋₁ antibody 15 (A₁, A₂) gave clear hybridization signals relative to the control antibody extract (A₃, A₄) and to control yeast RNA (B₁, B₂). Standards consisting of 1pg, 5pg, and 10pg of the purified, clone 81 cDNA fragment are shown in C1-3, respectively.

20 These results demonstrate that the particles captured from NANBH plasma by anti-NANB₅₋₁₋₁-antibodies contain nucleic acids which hybridize with HCV cDNA in clone 81, and thus provide further evidence that the cDNAs in these clones are derived from the etiologic agent for 25 NANBH.

IV.G. Immunological Reactivity of C100-3 with Purified Anti-NANB₅₋₁₋₁ Antibodies

The immunological reactivity of C100-3 fusion 30 polypeptide with anti-NANB₅₋₁₋₁ antibodies was determined by a radioimmunoassay, in which the antigens which were bound to a solid phase were challenged with purified anti-NANB₅₋₁₋₁ antibodies, and the antigen-antibody complex detected with ¹²⁵I-labeled sheep anti-human antibodies.

The immunological reactivity of C100-3 polypeptide was compared with that of SOD-NANB₅₋₁₋₁ antigen.

The fusion polypeptide C100-3 was synthesized and purified as described in Section IV.B.5. and in Section IV.B.6., respectively. The fusion polypeptide SOD-
5 NANB₅₋₁₋₁ was synthesized and purified as described in Section IV.B.1. and in Section IV.D.1., respectively. Purified anti-NANB₅₋₁₋₁ antibodies were obtained as described in Section IV.E.

One hundred microliter aliquots containing
10 varying amounts of purified C100-3 antigen in 0.125M Na borate buffer, pH 8.3, 0.075M NaCl (BBS) was added to each well of a microtiter plate (Dynatech Immulon 2 Removawell Strips). The plate was incubated at 4°C overnight in a humid chamber, after which, the protein solution was
15 removed and the wells washed 3 times with BBS containing 0.02% Triton X-100 (BBST). To prevent non-specific binding, the wells were coated with BSA by addition of 100 microliters of a 5 mg/ml solution of BSA in BBS followed by incubation at room temperature for 1 hour, after which
20 the excess BSA solution was removed. The polypeptides in the coated wells were reacted with purified anti-NANB₅₋₁₋₁ antibodies by adding 1 microgram antibody/well, and incubating the samples for 1 hr at 37°C. After incubation, the excess solution was removed by aspiration, and
25 the wells were washed 5 times with BBST. Anti-NANB₅₋₁₋₁ bound to the fusion polypeptides was determined by the binding of ¹²⁵I-labeled F'(ab)₂ sheep anti-human IgG to the coated wells. Aliquots of 100 microliters of the labeled probe (specific activity 5-20 microcuries/
30 microgram) were added to each well, and the plates were incubated at 37°C for 1 hour, followed by removal of excess probe by aspiration, and 5 washes with BBST. The amount of radioactivity bound in each well was determined by counting in a counter which detects gamma radiation.
35

The results of the immunological reactivity of C100 with purified anti-NANB₅₋₁₋₁ as compared to that of NANB₅₋₁₋₁ with the purified antibodies are shown in Table 3.

5

Table 3
Immunological Reactivity of C100-3 compared to NANB₅₋₁₋₁
by Radioimmunoassay

10	AG(ng)	RIA (cpm/assay)					
		400	320	240	160	60	0
	NANB ₅₋₁₋₁	7332	6732	4954	4050	3051	57
15	C100-3	7450	6985	5920	5593	4096	67

The results in Table 3 show that anti-NANB₅₋₁₋₁ recognizes an epitope(s) in the C100 moiety of the C100-3 polypeptide. Thus NANB₅₋₁₋₁ and C100 share a common 20 epitope(s). The results suggest that the cDNA sequence encoding this NANBV epitope(s) is one which is present in both clone 5-1-1 and in clone 81.

IV.H. Characterization of HCV
25

IV.H.1. Characterization of the Strandedness of the HCV Genome.

The HCV genome was characterized with respect to its strandedness by isolating the nucleic acid fraction 30 from particles captured on anti-NANB₅₋₁₋₁ antibody coated polystyrene beads, and determining whether the isolated nucleic acid hybridized with plus and/or minus strands of HCV cDNA.

Particles were captured from HCV infected 35 chimpanzee plasma using polystyrene beads coated with

immunopurified anti-NANB₅₋₁₋₁ antibody as described in Section IV.F.1. The nucleic acid component of the particles was released using the method described in Section IV.F.2. Aliquots of the isolated genomic nucleic acid equivalent to 3 mls of high titer plasma were blotted onto nitrocellulose filters. As controls, aliquots of denatured HCV cDNA from clone 81 (2 picograms) was also blotted onto the same filters. The filters were probed with ³²P-labeled mixture of plus or mixture of minus strands of single stranded DNA cloned from HCV cDNAs; the cDNAs were excised from clones 40b, 81, and 25c.

The single stranded probes were obtained by excising the HCV cDNAs from clones 81, 40b, and 25c with EcoRI, and cloning the cDNA fragments in M13 vectors, mp18 and mp19 [Messing (1983)]. The M13 clones were sequenced to determine whether they contained the plus or minus strands of DNA derived from the HCV cDNAs. Sequencing was by the dideoxychain termination method of Sanger et al. (1977).

Each of a set of duplicate filters containing aliquots of the HCV genome isolated from the captured particles was hybridized with either plus or minus strand probes derived from the HCV cDNAs. Fig. 41 shows the autoradiographs obtained from probing the NANBV genome with the mixture of probes derived from clones 81, 40b, and 25c. This mixture was used to increase the sensitivity of the hybridization assay. The samples in panel I were hybridized with the plus strand probe mixture. The samples in panel II were probed by hybridization with the minus strand probe mixture. The composition of the samples in the panels of the immunoblot are presented in table 4.

Table 4

lane		A	B
5	1	HCV genome	*
	2	----	*
10	3	*	cDNA 81
	4	----	cDNA 81

15 * is an undescribed sample.

As seen from the results in Fig. 41, only the minus strand DNA probe hybridizes with the isolated HCV genome. This result, in combination with the result showing that the genome is sensitive to RNase and not DNase (See Section IV.C.2.), suggests that the genome of NANBV is positive stranded RNA.

These data, and data from other laboratories concerning the physicochemical properties of a putative NANBV(s), are consistent with the possibility that HCV is a member of the Flaviviridae. However, the possibility that HCV represents a new class of viral agent has not been eliminated.

30 IV.H.2. Detection of Sequences in Captured Particles
Which When Amplified by PCR Hybridize to HCV cDNA Derived
from Clone 81

The RNA in captured particles was obtained as described in Section IV.H.1. The analysis for sequences 35 which hybridize to the HCV cDNA derived from clone 81 was

carried out utilizing the PCR amplification procedure, as described in Section IV.C.3, except that the hybridization probe was a kinased oligonucleotide derived from the clone 81 cDNA sequence. The results showed that the amplified 5 sequences hybridized with the clone 81 derived HCV cDNA probe.

IV.H.3. Homology Between the Non-Structural Protein of Dengue Flavivirus (MNWWVD1) and the HCV Polypeptides
10 Encoded by the Combined ORF of Clones 14i Through 39c

The combined HCV cDNAs of clones 14i through 39c contain one continuous ORF, as shown in Fig. 26. The polypeptide encoded therein was analyzed for sequence homology with the region of the non-structural 15 polypeptide(s) in Dengue flavivirus (MNWVD1). The analysis used the Dayhoff protein data base, and was performed on a computer. The results are shown in Fig. 42, where the symbol (:) indicates an exact homology, and the symbol (.) indicates a conservative replacement in the 20 sequence; the dashes indicate spaces inserted into the sequence to achieve the greatest homologies. As seen from the figure, there is significant homology between the sequence encoded in the HCV cDNA, and the non-structural polypeptide(s) of Dengue virus. In addition to the homology shown in Fig. 42, analysis of the polypeptide segment 25 encoded in a region towards the 3'-end of the cDNA also contained sequences which are homologous to sequences in the Dengue polymerase. Of consequence is the finding that the canonical Gly-Asp-Asp (GDD) sequence thought to be 30 essential for RNA-dependent RNA polymerases is contained in the polypeptide encoded in HCV cDNA, in a location which is consistent with that in Dengue 2 virus. (Data not shown.)

IV.H.4. HCV-DNA is Not Detectable in NANBH Infected Tissue

Two types of studies provide results suggesting that HCV-DNA is not detectable in tissue from an individual with NANBH. These results, in conjunction with those described in IV.C. and IV.H.1. and IV.H.2. provide evidence that HCV is not a DNA containing virus, and that its replication does not involve cDNA.

IV.H.4.a. Southern Blotting Procedure

In order to determine whether NANBH infected chimpanzee liver contains detectable HCV-DNA (or HCV-cDNA), restriction enzyme fragments of DNA isolated from this source was Southern blotted, and the blots probed with ³²P-labeled HCV cDNA. The results showed that the labeled HCV cDNA did not hybridize to the blotted DNA from the infected chimpanzee liver. It also did not hybridize to control blotted DNA from normal chimpanzee liver. In contrast, in a positive control, a labeled probe of the beta-interferon gene hybridized strongly to Southern blots of restriction enzyme digested human placental DNA. These systems were designed to detect a single copy of the gene which was to be detected with the labeled probe.

DNAs were isolated from the livers of two chimpanzees with NANBH. Control DNAs were isolated from uninjected chimpanzee liver, and from human placentas. The procedure for extracting DNA was essentially according to Maniatis et al. (1982), and the DNA samples were treated with RNase during the isolation procedure.

Each DNA sample was treated with either EcoRI, MboI, or HincII (12 micrograms), according to the manufacturer's directions. The digested DNAs were electrophoresed on 1% neutral agarose gels, Southern blotted onto nitrocellulose, and the blotted material hybridized with the appropriate nick-translated probe cDNA (3×10^6 cpm/ml of hybridization mix). The DNA from

infected chimpanzee liver and normal liver were hybridized with 32 P-labeled HCV cDNA from clones 36 plus 81; the DNA from human placenta was hybridized with 32 P-labeled DNA from the beta-interferon gene. After hybridization, the 5 blots were washed under stringent conditions, i.e., with a solution containing 0.1 x SSC, 0.1% SDS, at 65°C.

The beta-interferon gene DNA was prepared as described by Houghton et al (1981).

10 IV.H.4.b. Amplification by the PCR Technique

In order to determine whether HCV-DNA could be detected in liver from chimpanzees with NANBH, DNA was isolated from the tissue, and subjected to the PCR amplification-detection technique using primers and probe

15 Polynucleotides derived from HCV cDNA from clone 81.

Negative controls were DNA samples isolated from uninfected HepG2 tissue culture cells, and from presumably uninfected human placenta. Positive controls were samples of the negative control DNAs to which a known relatively 20 small amount (250 molecules) of the HCV cDNA insert from clone 81 was added.

In addition, to confirm that RNA fractions isolated from the same livers of chimpanzees with NANBH contained sequences complementary to the HCV-cDNA probe, 25 the PCR amplification-detection system was also used on the isolated RNA samples.

In the studies, the DNAs were isolated by the procedure described in Section IV.H.4.a, and RNAs were extracted essentially as described by Chirgwin et al. 30 (1981).

Samples of DNA were isolated from 2 infected chimpanzee livers, from uninfected HepG2 cells, and from human placenta. One microgram of each DNA was digested with HindIII according to the manufacturer's directions.

35 The digested samples were subjected to PCR amplification

and detection for amplified HCV cDNA essentially as described in Section IV.C.3., except that the reverse transcriptase step was omitted. The PCR primers and probe were from HCV cDNA clone 81, and are described in Section 5 IV.C.3.. Prior to the amplification, for positive controls, a one microgram sample of each DNA was "spiked" by the addition of 250 molecules of HCV cDNA insert isolated from clone 81.

In order to determine whether HCV sequences were 10 present in RNA isolated from the livers of chimpanzees with NANBH, samples containing 0.4 micrograms of total RNA were subjected to the amplification procedure essentially as described in Section IV.C.3., except that the reverse transcriptase was omitted from some of the samples as a 15 negative control. The PCR primers and probe were from HCV cDNA clone 81, as described supra.

The results showed that amplified sequences complementary to the HCV cDNA probe were not detectable in the DNAs from infected chimpanzee liver, nor were they 20 detectable in the negative controls. In contrast, when the samples, including the DNA from infected chimpanzee liver, was spiked with the HCV cDNA prior to amplification, the clone 81 sequences were detected in all positive control samples. In addition, in the RNA 25 studies, amplified HCV cDNA clone 81 sequences were detected only when reverse transcriptase was used, suggesting strongly that the results were not due to a DNA contamination.

These results show that hepatocytes from 30 chimpanzees with NANBH contain no, or undetectable levels, of HCV DNA. Based upon the spiking study, if HCV DNA is present, it is at a level far below .06 copies per hepatocyte. In contrast, the HCV sequences in total RNA from the same liver samples was readily detected with the 35 PCR technique.

IV.I. ELISA Determinations for HCV Infection Using HCV c100-3 As Test Antigen

All samples were assayed using the HCV c100-3 ELISA. This assay utilizes the HCV c100-3 antigen (which 5 was synthesized and purified as described in Section IV.B.5), and a horseradish peroxidase (HRP) conjugate of mouse monoclonal anti-human IgG.

Plates coated with the HCV c100-3 antigen were prepared as follows. A solution containing Coating buffer 10 (50mM Na Borate, pH 9.0), 21 ml/plate, BSA (25 micrograms/ml), c100-3 (2.50 micrograms/ml) was prepared just prior to addition to the Removeawell Immulon I plates (Dynatech Corp.). After mixing for 5 minutes, 0.2ml/well of the solution was added to the plates, they were covered and 15 incubated for 2 hours at 37°C, after which the solution was removed by aspiration. The wells were washed once with 400 microliters Wash Buffer (100 mM sodium phosphate, pH 7.4, 140 mM sodium chloride, 0.1% (W/V) casein, 1% (W/V) Triton x-100, 0.01% (W/V) Thimerosal). After removal 20 of the wash solution, 200 microliters/well of Postcoat solution (10 mM sodium phosphate, pH 7.2, 150 mM sodium chloride, 0.1% (w/v) casein and 2 mM phenylmethylsulfonylfluoride (PMSF)) was added, the plates were loosely covered to prevent evaporation, and were al- 25 lowed to stand at room temperature for 30 minutes. The wells were then aspirated to remove the solution, and lyophilized dry overnight, without shelf heating. The prepared plates may be stored at 2-8°C in sealed aluminum pouches.

In order to perform the ELISA determination, 30 20 microliters of serum sample or control sample was added to a well containing 200 microliters of sample diluent (100 mM sodium phosphate, pH 7.4, 500 mM sodium chloride, 1 mM EDTA, 0.1% (W/V) Casein, 0.015 (W/V) Therosal, 1% (W/V) 35 Triton X-100, 100 micrograms/ml yeast extract). The

plates were sealed, and incubated at 37°C for two hours, after which the solution was removed by aspiration, and the wells were washed with 400 microliters of wash buffer (phosphate buffered saline (PBS) containing 0.05% Tween 5 20). The washed wells were treated with 200 microliters of mouse anti-human IgG-HRP conjugate contained in a solution of Ortho conjugate diluent (10 mM sodium phosphate, pH 7.2, 150 mM sodium chloride, 50% (V/V) fetal bovine serum, 1% (V/V) heat treated horse serum, 1 mM K₃Fe(CN)₆, 10 0.05% (W/V) Tween 20, 0.02% (W/V) Thimerosal). Treatment was for 1 hour at 37°C, the solution was removed by aspiration, and the wells were washed with wash buffer, which was also removed by aspiration. To determine the amount of bound enzyme conjugate, 200 microliters of 15 substrate solution (10 mg O-phenylenediamine dihydrochloride per 5 ml of Developer solution) was added. Developer solution contains 50 mM sodium citrate adjusted to pH 5.1 with phosphoric acid, and 0.6 microliters/ml of 30% H₂O₂. The plates containing the substrate solution 20 were incubated in the dark for 30 minutes at room temperature, the reactions were stopped by the addition of 50 microliters/ml 4N sulfuric acid, and the ODs determined.

The examples provided below show that the 25 microtiter plate screening ELISA which utilizes HCV c100-3 antigen has a high degree of specificity, as evidenced by an initial rate of reactivity of about 1%, with a repeat reactive rate of about 0.5% on random donors. The assay is capable of detecting an immunoresponse in both the post 30 acute phase of the infection, and during the chronic phase of the disease. In addition, the assay is capable of detecting some samples which score negative in the surrogate tests for NANBH; these samples come from individuals with a history of NANBH, or from donors 35 implicated in NANBH transmission.

In the examples described below, the following abbreviations are used:

	ALT	Alanine amino transferase
5	Anti-HBc	Antibody against HBc
	Anti-HBsAg	Antibody against HBsAg
	HBc	Hepatitis B core antigen
	ABsAg	Hepatitis B surface antigen
	IgG	Immunoglobulin G
10	IgM	Immunoglobulin M
	IU/L	International units/Liter
	NA	Not available
	NT	Not tested
	N	Sample size
15	Neg	Negative
	OD	Optical density
	Pos	Positive
	S/CO	Signal/cutoff
	SD	Standard deviation
20	x	Average or mean
	WNL	Within normal limits

IV.I.1. HCV Infection in a Population of Random Blood

Donors

25 A group of 1,056 samples (fresh sera) from random blood donors were obtained from Irwin Memorial Blood Bank, San Francisco, California. The test results obtained with these samples are summarized in a histogram showing the distribution of the OD values (Fig. 43). As 30 seen in Fig. 43, 4 samples read >3, 1 sample reads between 1 and 3, 5 samples read between 0.4 and 1, and the remaining 1,046 samples read <0.4, with over 90% of these samples reading <0.1.

The results on the reactive random samples are 35 presented in Table 5. Using a cut-off value equal to the

mean plus 5 standard deviations, ten samples out of the 1,056 (0.95%) were initially reactive. Of these, five samples (0.47%) repeated as reactive when they were assayed a second time using the ELISA. Table 5 also shows 5 the ALT and Anti-HBd status for each of the repeatedly reactive samples. Of particular interest is the fact that all five repeat reactive samples were negative in both surrogate tests for NANBH, while scoring positive in the HCV ELISA.

10

15

20

25

30

35

TABLE 5
RESULTS ON REACTIVE RANDOM SAMPLES

5

$N = 1051$
 $\bar{x} = 0.049^*$
 $SD = \pm 0.074$
Cut-cff: $\bar{x} + 5SD = 0.419$ (0.400 + Negative Control)

10

		Initial Reactives	Repeat Reactives	Anti HBc***
	Samples	OD	OD	ALT** (IU/L)
15	4227	0.462	0.084	NA
	6292	0.569	0.294	NA
	6188	0.699	0.326	NA
	6157	0.735	0.187	NA
	6277	0.883	0.152	NA
	6397	1.567	1.392	30.14
	6019	>3.000	>3.000	46.48
20	6651	>3.000	>3.000	48.53
	6669	>3.000	>3.000	60.53
	4003	>3.000	3.000	WNL****
		10/1056 = 0.95%	5/1056 = 0.47%	Negative

* Samples reading >1.5 were not included in calculating the Mean and SD

** ALT \geq 68 IU/L is above normal limits.

25 *** Anti-HBc \leq 0.535 (competition assay) is considered positive.

**** WNL: Within normal limits.

30

35

IV.I.2. Chimpanzee Serum Samples

Serum samples from eleven chimpanzees were tested with the HCV c100-3 ELISA. Four of these 5 chimpanzees were infected with NANBH from a contaminated batch of Factor VIII (presumably Hutchinson strain), following an established procedure in a collaboration with Dr. Daniel Bradley at the Centers for Disease Control. As controls, four other chimpanzees were infected with HAV 10 and three with HBV. Serum samples were obtained at different times after infection.

The results, which are summarized in Table 6, show documented antibody seroconversion in all chimpanzees infected with the Hutchinson strain of NANBH. Following 15 the acute phase of infection (as evidenced by the significant rise and subsequent return to normal of ALT levels), antibodies to HCV c100-3 became detectable in the sera of the 4/4 NANBH infected chimpanzees. These samples had previously been shown, as discussed in Section 20 IV.B.3., to be positive by a Western analysis, and an RIA. In contrast, none of the control chimpanzees which had been infected with HAV or HBV showed evidence of reactivity in the ELISA.

25

30

35

TABLE 6
CHIMPANZEE SERUM SAMPLES

		<u>OD</u>	<u>S/CO</u>	<u>INOCULATION DATE</u>	<u>BLEED DATE</u>	<u>ALT (IU/L)</u>	<u>TRANSFUSED</u>
5							
	NEGATIVE CONTROL	0.001					
	POSITIVE CONTROL	1.501					
	CUTOFF	0.401					
10	<u>Chimp 1</u>	-0.007	0.00	05/24/84	05/24/84	9	NANB
		0.003	0.01		08/07/84	71	
		>3.000	>7.48		09/18/84	19	
		>3.000	>7.48		10/21/84	---	
15	<u>Chimp 2</u>	---	---	06/07/84	---	---	NANB
		-0.003	0.00		05/31/84	5	
		-0.005	0.00		06/28/84	52	
		0.945	2.36		08/20/84	13	
20	<u>Chimp 3</u>	>3.000	>7.48		10/24/84	---	
		0.005	0.01	03/14/85	03/14/85	8	NANB
		0.017	0.04		04/26/85	205	
		0.006	0.01		05/06/85	14	
25	<u>Chimp 4</u>	1.010	2.52		08/20/85	6	
		-0.006	0.00	03/11/85	03/11/85	11	NANB
		0.003	0.01		05/09/85	132	
		0.523	1.31		06/06/85	---	
30	<u>Chimp 5</u>	1.571	3.93		08/01/85	---	
		-0.006	0.00	11/21/80	11/21/80	4	HAV
		0.001	0.00		12/16/80	147	
		0.003	0.01		12/30/80	18	
35	<u>Chimp 6</u>	0.006	0.01		07/29 - 08/21/81	5	
		---	---	05/25/82	---	---	HAV
		-0.005	0.00		05/17/82	---	
		0.001	0.00		06/10/82	106	
40	<u>Chimp 7</u>	-0.004	0.00		07/06/82	10	
		0.290	0.72		10/01/82	---	
		-0.008	0.00	05/25/82	05/25/82	7	HAV
		-0.004	0.00		06/17/82	83	
45	<u>Chimp 8</u>	-0.006	0.00		09/16/82	5	
		0.005	0.01		10/09/82	---	

TABLE 6

CHIMPANZEE SERUM SAMPLES

(Cont'd)

		OD	S/CO	INOCULATION DATE	BLEED DATE	ALT (IU/L)	TRANSFUSED
5	<u>Chimp 8</u>	-0.007	0.00	11/21/80	11/21/80	15	HAV
		0.000	0.00		12/16/80	130	
		0.004	0.01		02/03/81	8	
		0.000	0.00		06/03 - 06/10/81	4.5	
10	<u>Chimp 9</u>	---	---	07/24/80	---	---	HBV
		0.019	0.05		08/22 - 10/10/79	---	
		---	---		03/11/81	57	
		0.015	0.04		07/01 - 08/05/81	9	
		0.008	0.02		10/01/81	6	
15	<u>Chimp 10</u>	---	---	05/12/82	---	---	HDV
		0.011	0.03		04/21 - 05/12/82	9	
		0.015	0.04		09/01 - 09/08/82	126	
		0.008	0.02		12/02/82	9	
		0.010	0.02		01/06/83	13	
20	<u>Chimp 11</u>	---	---	05/12/82	---	---	HDV
		0.000	0.00		01/06 - 05/12/82	11	
		---	---		06/23/82	100	
		-0.003	0.00		06/09 - 07/07/82	---	
		-0.003	0.00		10/28/82	9	
		-0.003	0.00		12/20/82	10	

25

30

35

IV.I.3. Panel 1: Proven Infectious Sera from Chronic Human NANBH Carriers

A coded panel consisted of 22 unique samples, each one in duplicate, for a total of 44 samples. The 5 samples were from proven infectious sera from chronic NANBH carriers, infectious sera from implicated donors, and infectious sera from acute phase NANBH patients. In addition, the samples were from highly pedigreed negative controls, and other disease controls. This panel was 10 provided by Dr. H. Alter of the Department of Health and Human Services, National Institutes of Health, Bethesda, Maryland. The panel was constructed by Dr. Alter several years ago, and has been used by Dr. Alter as a qualifying panel for putative NANBH assays.

15 The entire panel was assayed twice with the ELISA assay, and the results were sent to Dr. Alter to be scored. The results of the scoring are shown in Table 7. Although the Table reports the results of only one set of duplicates, the same values were obtained for each of the 20 duplicate samples.

As shown in Table 7, 6 sera which were proven infectious in a chimpanzee model were strongly positive. The seventh infectious serum corresponded to a sample for an acute NANBH case, and was not reactive in this ELISA. 25 A sample from an implicated donor with both normal ALT levels and equivocal results in the chimpanzee studies was non-reactive in the assay. Three other serial samples from one individual with acute NANBH were also non-reactive. All samples coming from the highly pedigreed 30 negative controls, obtained from donors who had at least 10 blood donations without hepatitis implication, were non-reactive in the ELISA. Finally, four of the samples tested had previously scored as positive in putative NANBH assays developed by others, but these assays were not

confirmable. These four samples scored negatively with the HCV ELISA.

5

10

15

20

25

30

35

TABLE 7
H. ALTER'S PANEL 1:

5

	PANEL	1ST RESULT	2ND RESULT
	1) PROVEN INFECTIOUS BY CHIMPANZEE TRANSMISSION		
	A. CIRRHOTIC NANB; Post-Tx		
10	JF	+	+
	ED	+	+
	PG	+	+
	B. IMPLICATED DONORS WITH Elevated ALT		
	BC	+	+
	JJ	+	+
	BB	+	+
	C. ACUTE NANB; Post-Tx		
15	WII	-	-
	2) EQUIVOCALLY INFECTIOUS BY CHIMPANZEE TRANSMISSION		
	A. IMPLICATED DONOR WITH NORMAL ALT		
	CC	-	-
	3) ACUTE NANB; Post-Tx		
20	JL WEEK 1	-	-
	JL WEEK 2	-	-
	JL WEEK 3	-	-
	4) DISEASE CONTROLS		
	A. PRIMARY BILIARY CIRRHOSIS		
	EK	-	-
	B. ALCOHOLIC HEPATITIS IN RECOVERY		
	IIB	-	-
	5) PEDIGREADED NEGATIVE CONTROLS		
25	DM	-	-
	DC	-	-
	LV	-	-
	HL	-	-
	AII	-	-
	6) POTENTIAL NANB "ANTIGENS"		
	JS-80-DIT-O (ISHIDA)	-	-
	ASTERIX (TREPO)	-	-
30	ZURTZ (ARNOLD)	-	-
	BECASSDINE (TREPO)	-	-

35

IV.I.4. Panel 2: Donor/Recipient NANBH

The coded panel consisted of 10 unequivocal donor-recipient cases of transfusion associated NANBH, 5 with a total of 188 samples. Each case consisted of samples of some or all the donors to the recipient, and of serial samples (drawn 3, 6, and 12 months after transfusion) from the recipient. Also included was a pre-bleed, drawn from the recipient before transfusion. The 10 coded panel was provided by Dr. H. Alter, from the NIH, and the results were sent to him for scoring.

The results, which are summarized in Table 8, show that the ELISA detected antibody seroconversion in 9 of 10 cases of transfusion associated NANBH. Samples from 15 case 4 (where no seroconversion was detected), consistently reacted poorly in the ELISA. Two of the 10 recipient samples were reactive at 3 months post transfusion. At six months, 8 recipient samples were reactive; and at twelve months, with the exception of case 4, all 20 samples were reactive. In addition, at least one antibody positive donor was found in 7 out of the 10 cases, with case 10 having two positive donors. Also, in case 10, the recipient's pre-bleed was positive for HCV antibodies. The one month bleed from this recipient dropped to border- 25 line reactive levels, while it was elevated to positive at 4 and 10 month bleeds. Generally, a S/CO of 0.4 is considered positive. Thus, this case may represent a prior infection of the individual with HCV.

The ALT and HBc status for all the reactive, 30 i.e., positive, samples are summarized in Table 9. As seen in the table, 1/8 donor samples was negative for the surrogate markers and reactive in the HCV antibody ELISA. On the other hand, the recipient samples (followed up to 12 months after transfusion) had either elevated ALT, 35 positive Anti-HBc, or both.

TABLE 8
DONOR/RECIPIENT NANB PANEL

5

10

CASE	DONOR		RECIPIENT		POST-TX			
	OD	S/CO	PREBLEED	3 MONTHS	OD	S/CO	OD	S/CO
15	1.	---	.032	0.07	.112	0.26	>3.000	>6.96
	2.	---	.059	0.14	.050	0.12	1.681	3.90
	3.	.403	0.94	.049	0.11	.057	0.13	>3.000
	4.	---	---	.065	0.15	.073	0.17	.067
	5.	>3.000	>6.96	.034	0.08	.096	0.22	>3.000
20	6.	>3.000	>6.96	.056	0.13	1.475	3.44	>3.000
	7.	>3.000	>6.96	.034	0.08	.056	0.13	>3.000
	8.	>3.000	>6.96	.061	0.14	.078	0.18	2.262
	9.	>3.000	>6.96	.080	0.19	.127	0.30	.055
	10.	>3.000	>6.96	>3.000	>6.96	.317*	0.74	>3.000**
25		>3.000	>6.96					>3.000***

* 1 MONTH, ** 4 MONTHS, *** 10 MONTHS

30

35

TABLE 9

ALT AND HBc STATUS FOR REACTIVE SAMPLES IN
H. ALTER PANEL 1

		<u>Samples</u>	<u>Anti- ALT*</u>	<u>HBc**</u>
<u>Donors</u>				
10	Case 3		Normal	Negative
	Case 5		Elevated	Positive
	Case 6		Elevated	Positive
	Case 7		Not available	Negative
	Case 8		Normal	Positive
	Case 9		Elevated	Not available
	Case 10		Normal	Positive
	Case 10		Normal	Positive
15	<u>Recipients</u>			
	Case 1 12 mo	6 mo Elevated	Elevated Not tested	Positive
	Case 2 12 mo	6 mo Elevated	Elevated Not tested	Negative
	Case 3 12 mo	6 mo Elevated	Normal Not tested***	Not tested***
	Case 5 12 mo	6 mo Elevated	Elevated Not tested	Not tested
	Case 6 6 mo 12 mo	3 mo Elevated Elevated	Elevated Negative Not tested	Negative
	Case 7 12 mo	6 mo Elevated	Elevated Negative	Negative
	Case 8 12 mo	6 mo Elevated	Normal Not tested	Positive
30	Case 9	12 mo	Elevated	Not tested
	Case 10 10 mo	4 mo Elevated	Elevated Not tested	Not tested

* ALT \geq 45 IU/L is above normal limits.

** Anti-HBc \leq 50% (competition assay) is considered positive.

*** Prebleed and 3 mo samples were negative for HBc.

IV.I.5. Determination of HCV Infection in High Risk Group Samples

Samples from high risk groups were monitored
5 using the ELISA to determine reactivity to HCV c100-3 antigen. These samples were obtained from Dr. Gary Tegtmeier, Community Blood Bank, Kansas City. The results are summarized in Table 10.

As shown in the table, the samples with the
10 highest reactivity are obtained from hemophiliacs (76%). In addition, samples from individuals with elevated ALT and positive for Anti-HBc, scored 51% reactive, a value which is consistent with the value expected from clinical data and NANBH prevalence in this group. The incidence of
15 antibody to HCV was also higher in blood donors with elevated ALT alone, blood donors positive for antibodies to Hepatitis B core alone, and in blood donors rejected for reasons other than high ALT or anti-core antibody when compared to random volunteer donors.

20

25

30

35

TABLE 10

NANBH HIGH RISK GROUP SAMPLES

5

	Group	N	Distribution		% Reactive
			N	OD	
10	Elevated ALT	35	3	>3.000	11.4%
		1 0.728			
10	Anti-HBc	24	5	>3.000	20.8%
10	Elevated ALT, Anti-HBc	33	12	>3.000	51.5%
		1 2.768			
		1 2.324			
15		1 0.939			
15		1 0.951			
15		1 0.906			
15	Rejected Donors	25	5	>3.000	20.0%
20	Donors with History of Hepatitis	150	19	>3.000	14.7%
20		1 0.837			
20		1 0.714			
20		1 0.469			
25	Haemophiliacs	50	31	>3.000	76.0%
25		1 2.568			
25		1 2.483			
25		1 2.000			
25		1 1.979			
25		1 1.495			
25		1 1.209			
25		1 0.819			

30

35

IV.I.6 Comparative Studies Using Anti-IgG or Anti-IgM Monoclonal Antibodies, or Polyclonal Antibodies as a Second Antibody in the HCV c100-3 ELISA

5 The sensitivity of the ELISA determination which uses the anti-IgG monoclonal conjugate was compared to that obtained by using either an anti-IgM monoclonal conjugate, or by replacing both with a polyclonal antiserum reported to be both heavy and light chain specific. The following studies were performed.

10 IV.I.6.a. Serial Samples from Seroconverters
15 Serial samples from three cases of NANB seroconverters were studied in the HCV c100-3 ELISA assay using in the enzyme conjugate either the anti-IgG monoclonal alone, or in combination with an anti-IgM monoclonal, or using a polyclonal antiserum. The samples were provided by Dr. Cladd Stevens, N.Y. Blood Center, N.Y.C., N.Y.. The sample histories are shown in Table 11.

20 The results obtained using an anti-IgG monoclonal antibody-enzyme conjugate are shown in Table 12. The data shows that strong reactivity is initially detected in samples 1-4, 2-8, and 3-5, of cases 1, 2, and 3, respectively.

25 The results obtained using a combination of an anti-IgG monoclonal conjugate and an anti-IgM conjugate are shown in Table 13. Three different ratios of anti-IgG to anti-IgM were tested; the 1:10,000 dilution of anti-IgG was constant throughout. Dilutions tested for the anti-IgM monoclonal conjugate were 1:30,000, 1:60,000, and 1:120,000. The data shows that, in agreement with the studies with anti-IgG alone, initial strong reactivity is detected in samples 1-4, 2-8, and 3-5.

30 The results obtained with the ELISA using anti-IgG monoclonal conjugate (1:10,000 dilution), or Tago

polyclonal conjugate (1:80,000 dilution), or Jackson polyclonal conjugate (1:80,000 dilution) are shown in Table 14. The data indicates that initial strong reactivity is detected in samples 1-4, 2-8, and 3-5 using all 5 three configurations; the Tago polyclonal antibodies yielded the lowest signals.

The results presented above show that all three configurations detect reactive samples at the same time after the acute phase of the disease (as evidenced by the 10 ALT elevation). Moreover, the results indicate that the sensitivity of the HCV c100-3 ELISA using anti-IgG monoclonal-enzyme conjugate is equal to or better than that obtained using the other tested configurations for the enzyme conjugate.

15

20

25

30

35

TABLE 11
DESCRIPTION OF SAMPLES FROM CLADD STEVENS PANEL

5

	Date	HBsAg	Anti-HBs	Anti-HBc	ALT	Bilirubin
--	------	-------	----------	----------	-----	-----------

Case 1

10	1-1	8/5/81	1.0	91.7	12.9	40.0	-1.0
	1-2	9/2/81	1.0	121.0	15.1	274.0	1.4
	1-3	10/7/81	1.0	64.0	23.8	261.0	0.9
	1-4	11/19/81	1.0	67.3	33.8	75.0	0.9
	1-5	12/15/81	1.0	50.5	27.6	71.0	1.0

Case 2

15	2-1	10/19/81	1.0	1.0	116.2	17.0	-1.0	
	2-2	11/17/81	1.0	0.8	89.5	46.0	1.1	
	2-3	12/02/81	1.0	1.2	78.3	63.0	1.4	
	2-4	12/14/81	1.0	0.9	90.6	152.0	1.4	
	2-5	12/23/81	1.0	0.8	93.6	624.0	1.7	
	20	2-6	1/20/82	1.0	0.8	92.9	66.0	1.5
		2-7	2/15/82	1.0	0.8	86.7	70.0	1.3
		2-8	3/17/82	1.0	0.9	69.8	24.0	-1.0
		2-9	4/21/82	1.0	0.9	67.1	53.0	1.5
		2-10	5/19/82	1.0	0.5	74.8	95.0	1.6
		2-11	6/14/82	1.0	0.8	82.9	37.0	-1.0

25

Case 3

30	3-1	4/7/81	1.0	1.2	88.4	13.0	-1.0
	3-2	5/12/81	1.0	1.1	126.2	236.0	0.4
	3-3	5/30/81	1.0	0.7	99.9	471.0	0.2
	3-4	6/9/81	1.0	1.2	110.8	315.0	0.4
	3-5	7/6/81	1.0	1.1	89.9	273.0	0.4
	3-6	8/10/81	1.0	1.0	118.2	158.0	0.4
	3-7	9/8/81	1.0	1.0	112.3	84.0	0.3
	3-8	10/14/81	1.0	0.9	102.5	180.0	0.5
	3-9	11/11/81	1.0	1.0	84.6	154.0	0.3

35

TABLE 12

ELISA RESULTS OBTAINED USING AN ANTI-IgG
MONOCLOINAL CONJUGATE

5

	<u>SAMPLE</u>	<u>DATE</u>	<u>ALT</u>	<u>OD</u>	<u>S/CO</u>
10	NEG CONTROL			.076	
	CUTOFF			.476	
	PC (1:128)			1.390	
15	<u>CASE #1</u>				
	1-1	08/05/81	40.0	.178	.37
	1-2	09/02/81	274.0	.154	.32
	1-3	10/07/81	261.0	.129	.27
	1-4	11/19/81	75.0	.937	1.97
	1-5	12/15/81	71.0	>3.000	>6.30
20	<u>CASE #2</u>				
	2-1	10/19/81	17.0	.058	0.12
	2-2	11/17/81	46.0	.050	0.11
	2-3	12/02/81	63.0	.047	0.10
	2-4	12/14/81	152.0	.059	0.12
	2-5	12/23/81	624.0	.070	0.15
	2-6	01/20/82	66.0	.051	0.11
	2-7	02/15/82	70.0	.139	0.29
	2-8	03/17/82	24.0	1.867	3.92
	2-9	04/21/82	53.0	>3.000	>6.30
	2-10	05/19/82	95.0	>3.000	>6.30
25	2-11	06/14/82	37.0	>3.000	>6.30
	<u>CASE #3</u>				
	3-1	04/07/81	13.0	.090	.19
	3-2	05/12/81	236.0	.064	.13
	3-3	05/30/81	471.0	.079	.17
	3-4	06/09/81	315.0	.211	.44
	3-5	07/06/81	273.0	1.707	3.59
	3-6	08/10/81	158.0	>3.000	>6.30
	3-7	09/08/81	84.0	>3.000	>6.30
30	3-8	10/14/81	180.0	>3.000	>6.30
	3-9	11/11/81	154.0	>3.000	>6.30
	35				

TABLE 13
ELISA RESULTS OBTAINED USING ANTI-IgG and ANTI-IgM
MONOCLONAL CONJUGATE

5	<u>SAMPLE</u>	<u>DATE</u>	<u>ALT</u>	<u>NANB ELISAS</u>		<u>MONOCLONALS</u>		<u>MONOCLONALS</u>	
				<u>MONOCLONALS</u>		<u>MONOCLONALS</u>		<u>MONOCLONALS</u>	
				IgG 1:10K	IgG 1:10K	IgG 1:10K	IgM 1:30K	IgM 1:60K	IgM 1:120K
				OD	S/CO	OD	OD	S/CO	OD
10	NEG CONTROL			.100		.080			.079
	CUTOFF					1.328			1.197
	PC (1:128)			1.083					
	<u>CASE #1</u>								
15	1-1	08/05/81	40	.173		.162			.070
	1-2	09/02/81	274	.194		.141			.079
	1-3	10/07/81	261	.162		.129			.063
	1-4	11/19/81	75	.812		.85			.709
	1-5	12/15/81	71	>3.00		>3.00			>3.00
	<u>Case #2</u>								
20	2-1	10/19/81	17	.442		.045			.085
	2-2	11/17/81	46	.102		.029			.030
	2-3	12/02/81	63	.059		.036			.027
	2-4	12/14/81	152	.065		.041			.025
	2-5	12/23/81	624	.082		.033			.032
	2-6	01/20/82	66	.102		.042			.027
	2-7	02/15/82	70	.188		.068			.096
25	2-8	03/17/82	24	1.728		1.668			1.541
	2-9	04/21/82	53	>3.00		2.443			>3.00
	2-10	05/19/82	95	>3.00		>3.00			>3.00
	2-11	06/14/82	37	>3.00		>3.00			>3.00
	<u>CASE #3</u>								
30	3-1	04/07/81	13	.193		.076			.049
	3-2	05/12/81	236	.201		.051			.038
	3-3	05/30/81	471	.132		.067			.052
	3-4	06/09/81	315	.175		.155			.140
	3-5	07/06/81	273	1.335		1.238			1.260
	3-6	08/10/81	158	>3.00		>3.00			>3.00
	3-7	09/08/81	84	>3.00		>3.00			>3.00
35	3-8	10/14/81	180	>3.00		>3.00			>3.00
	3-9	11/11/81	154	>3.00		>3.00			>3.00

TABLE 14

ELISA RESULTS OBTAINED USING POLYCLONAL CONJUGATES

5

	SAMPLE	DATE	ALT	MONOCLONAL		TAGO		JACKSON	
				1:10K		1:80K		1:80K	
				OD	S/CO	OD	S/CO	OD	S/CO
10	NEG CONTROL				.076		.045		.154
	CUTOFF				.476		.545		.654
	PC (1:128)				<u>1.390</u>		<u>.727</u>		<u>2.154</u>
<u>CASE #1</u>									
15	1-1	08/05/81	40	.178	.37	.067	.12	.153	.23
	1-2	09/02/81	274	.154	.32	.097	.18	.225	.34
	1-3	10/07/81	261	.129	.27	.026	.05	.167	.26
	1-4	11/19/81	75	.937	1.97	.324	.60	.793	1.21
	1-5	12/15/81	71	>3.00	>6.30	1.778	3.27	>3.00	>4.59
<u>CASE #2</u>									
20	2-1	10/19/81	17	.058	.12	.023	.04	.052	.08
	2-2	11/17/81	46	.050	.11	.018	.03	.058	.09
	2-3	12/02/81	63	.047	.10	.020	.04	.060	.09
	2-4	12/14/81	152	.059	.12	.025	.05	.054	.08
	2-5	12/23/81	624	.070	.15	.026	.05	.074	.11
25	2-6	01/20/82	66	.051	.11	.018	.03	.058	.09
	2-7	02/15/82	70	.139	.29	.037	.07	.146	.22
	2-8	03/17/82	24	1.867	3.92	.355	.65	1.429	2.19
	2-9	04/21/82	53	>3.00	>6.30	.748	1.37	>3.00	>4.59
	2-10	05/19/82	95	>3.00	>6.30	1.025	1.88	>3.00	>4.59
30	2-11	06/14/82	37	>3.00	>6.30	.917	1.68	>3.00	>4.59
	<u>CASE #3</u>								
	3-1	04/07/81	13	.090	.19	.049	.09	.138	.21
	3-2	05/12/81	236	.064	.13	.040	.07	.094	.14
	3-3	05/30/81	471	.079	.17	.045	.08	.144	.22
35	3-4	06/09/81	315	.211	.44	.085	.16	.275	.42
	3-5	07/06/81	273	1.707	3.59	.272	.50	1.773	2.71
	3-6	08/10/81	158	>3.00	>6.30	1.347	2.47	>3.00	>4.59
	3-7	09/08/81	84	>3.00	>6.30	2.294	4.21	>3.00	>4.59
	3-8	10/14/81	180	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59
<u>CASE #4</u>									
40	4-1	11/11/81	154	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59
	4-2	12/15/81	154	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59
	4-3	01/12/82	154	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59
	4-4	02/15/82	154	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59
	4-5	03/17/82	154	>3.00	>6.30	>3.00	>5.50	>3.00	>4.59

IV.I.6.b. Samples from Random Blood Donors

Samples from random blood donors (See Section IV.I.1.) were screened for HCV infection using the HCV 5 c100-3 ELISA, in which the antibody-enzyme conjugate was either an anti-IgG monoclonal conjugate, or a polyclonal conjugate. The total number of samples screened were 1077 and 1056, for the polyclonal conjugate and the monoclonal conjugate, respectively. A summary of the results of the 10 screening is shown in Table 15, and the sample distributions are shown in the histogram in Fig. 44.

The calculation of the average and standard deviation was performed excluding samples that gave a signal over 1.5, i.e., 1073 OD values were used for the 15 calculations utilizing the polyclonal conjugate, and 1051 for the anti-IgG monoclonal conjugate. As seen in Table 15, when the polyclonal conjugate was used, the average was shifted from 0.0493 to 0.0931, and the standard deviation was increased from 0.074 to 0.0933. Moreover, 20 the results also show that if the criteria of $x + 5SD$ is employed to define the assay cutoff, the polyclonal-enzyme conjugate configuration in the ELISA requires a higher cutoff value. This indicates a reduced assay specificity as compared to the monoclonal system. In addition, as 25 depicted in the histogram in Fig. 44, a greater separation of results between negative and positive distributions occurs when random blood donors are screened in an ELISA using the anti-IgG monoclonal conjugate as compared to the assay using a commercial polyclonal label.

TABLE 15

COMPARISON OF TWO ELISA CONFIGURATIONS IN
TESTING SAMPLES FROM RANDOM BLOOD DONORS

10

<u>CONJUGATE</u>	<u>POLYCLONAL</u>	<u>ANTI-IgG MONOCLONAL</u>
	(Jackson)	
Number of samples	1073	1051
Average (\bar{x})	0.0931	0.04926
Standard deviation (SD)	0.0933	0.07427
5 SD	0.4666	0.3714
CUT-OFF ($5 \text{ SD} + \bar{x}$)	0.5596	0.4206

20

25

30

35

IV.J. Detection of HCV Seroconversion in NANBH Patients from a Variety of Geographical Locations

Sera from patients who were suspected to have
5 NANBH based upon elevated ALT levels, and who were
negative in HAV and HBV tests were screened using the RIA
essentially as described in Section IV.D., except that the
HCV C100-3 antigen was used as the screening antigen in
the microtiter plates. As seen from the results presented
10 in Table 16, the RIA detected positive samples in a high
percentage of the cases.

Table 16
Seroconversion Frequencies for Anti-c100-3
15 Among NANBH Patients in Different Countries

	<u>Country</u>	<u>The Netherlands</u>	<u>Italy</u>	<u>Japan</u>
	No.			
20	Examined	5	36	26
	No.			
	Positive	3	29	19
25	%			
	Positive	60	80	73

IV.K. Detection of HCV Seroconversion in Patients with "Community Acquired" NANBH

Sera which was obtained from 100 patients with
NANBH, for whom there was no obvious transmission route
(i.e., no transfusions, i.v. drug use, promiscuity, etc.
were identified as risk factors), was provided by Dr. M.
35 Alter of the Center for Disease Control, and Dr. J.

Dienstag of Harvard University. These samples were screened using an RIA essentially as described in Section IV.D., except that the HCV c100-3 antigen was used as the screening antigen attached to the microtiter plates. The 5 results showed that of the 100 serum samples, 55 contained antibodies that reacted immunologically with the HCV c100-3 antigen.

- The results described above suggest that "Community Acquired" NANBH is also caused by HCV.
10 Moreover, since it has been demonstrated herein that HCV is related to Flaviviruses, most of which are transmitted by arthropods, it is suggestive that HCV transmission in the "Community Acquired" cases also results from arthropod transmission.

15

IV.L. Comparison of Incidence of HCV Antibodies and Surrogate Markers in Donors Implicated in NANBH Transmission

A prospective study was carried out to determine 20 whether recipients of blood from suspected NANBH positive donors, who developed NANBH, seroconverted to anti-HCV- antibody positive. The blood donors were tested for the surrogate marker abnormalities which are currently used as markers for NANBH infection, i.e., elevated ALT levels, 25 and the presence of anti-core antibody. In addition, the donors were also tested for the presence of anti-HCV antibodies. The determination of the presence of anti-HCV antibodies was determined using a radioimmunoassay as described in Section IV.K. The results of the study are 30 presented in Table 17, which shows: the patient number (column 1); the presence of anti-HCV antibodies in patient serum (column 2); the number of donations received by the patient, with each donation being from a different donor (column 3); the presence of anti-HCV antibodies in donor 35 serum (column 4); and the surrogate abnormality of the

donor (column 5) (NT or -- means not tested) (ALT is elevated transaminase, and ANTI-HBc is anti-core antibody).

The results in Table 17 demonstrate that the HCV antibody test is more accurate in detecting infected blood donors than are the surrogate marker tests. Nine out of ten patients who developed NANBH symptoms tested positive for anti-HCV antibody seroconversion. Of the 11 suspected donors, (patient 6 received donations from two different individuals suspected of being NANBH carriers), 9 were positive for anti-HCV antibodies, and 1 was borderline positive, and therefore equivocal (donor for patient 1). In contrast, using the elevated ALT test 6 of the ten donors tested negative, and using the anticore-antibody test 5 of the ten donors tested negative. Of greater consequence, though, in three cases (donors to patients 8, 9, and 10) the ALT test and the ANTI-HBc test yielded inconsistent results.

20

25

30

35

Table 17
DEVELOPMENT OF ANTI-HCV ANTIBODIES IN PATIENTS RECEIVING BLOOD FROM DONORS SUSPECTED OF BEING NANHB CARRIERS

Patient	Anti-HCV Seroconversion in Patient	No. of Donations/Donors	Anti-HCV Positive Donor	Surrogate Abnormality Alt Anti-HB
1	yes	18	equiv	no no
2	yes	18	yes NT	yes
3	yes	13	yes	no no
4	no	18	no	-- --
5	yes	16	yes	yes yes
6	yes	11	yes(2)	no no
7	yes	15	yes	NT no
8	yes	20	yes	no yes
9	yes	5	yes	yes no
10	yes	15	yes	no yes

*Same donor as anti-NANBV positive.

IV.M. Amplification for Cloning of HCV cDNA Sequences
Utilizing the PCR and Primers Derived from Conserved
Regions of Flavivirus Genomic Sequences

The results presented supra., which suggest that
5 HCV is a flavivirus or flavi-like virus, allows a strategy
for cloning uncharacterized HCV cDNA sequences utilizing
the PCR technique, and primers derived from the regions
encoding conserved amino acid sequences in flaviviruses.
Generally, one of the primers is derived from a defined
10 HCV genomic sequence, and the other primer which flanks a
region of unsequenced HCV polynucleotide is derived from a
conserved region of the flavivirus genome. The flavivirus
genomes are known to contain conserved sequences within
the NS1, and E polypeptides, which are encoded in the 5'-
15 region of the flavivirus genome. Corresponding sequences
encoding these regions lie upstream of the HCV cDNA
sequence shown in Fig. 26. Thus, to isolate cDNA
sequences derived from this region of the HCV genome,
upstream primers are designed which are derived from the
20 conserved sequences within these flavivirus polypeptides.
The downstream primers are derived from an upstream end of
the known portion of the HCV cDNA.

Because of the degeneracy of the code, it is
probable that there will be mismatches between the
25 flavivirus probes and the corresponding HCV genomic
sequence. Therefore a strategy which is similar to the
one described by Lee (1988) is used. The Lee procedure
utilizes mixed oligonucleotide primers complementary to
the reverse translation products of an amino acid
30 sequence; the sequences in the mixed primers takes into
account every codon degeneracy for the conserved amino
acid sequence.

Three sets of primer mixes are generated, based
on the amino acid homologies found in several
35 flaviviruses, including Dengue-2,4 (D-2,4), Japanese

Encephalitis Virus (JEV), Yellow Fever (YF), and West Nile Virus (WN). The primer mixture derived from the most upstream conserved sequence (5'-1), is based upon the amino acid sequence gly-trp-gly, which is part of the 5 conserved sequence asp-arg-gly-trp-gly-aspN found in the E protein of D-2, JEV, YF, and WN. The next primer mixture (5'-2) is based upon a downstream conserved sequence in E protein, phe-asp-gly-asp-ser-tyr-ileu-phe-gly-asp-ser-tyr-ileu, and is derived from phe-gly-asp; the conserved 10 sequence is present in D-2, JEV, YF, and WN. The third primer mixture (5'-3), is based on the amino acid sequence arg-ser-cys, which is part of the conserved sequence cys-cys-arg-ser-cys in the NS1 protein of D-2, D-4, JEV, YF, and WN. The individual primers which form the mixture in 15 5'-3 are shown in Fig. 45. In addition to the varied sequences derived from conserved region, each primer in each mixture also contains a constant region at the 5'-end which contains a sequence encoding sites for restriction enzymes, HindIII, MboI, and EcoRI.

20 The downstream primer, ssc5h20A, is derived from a nucleotide sequence in clone 5h, which contains HCV cDNA with sequences with overlap those in clones 14i and 11b. The sequence of ssc5h20A is

25 5' GTA ATA TGG TGA CAG AGT CA 3'.

An alternative primer, ssc5h34A, may also be used. This primer is derived from a sequence in clone 5h, and in addition contains nucleotides at the 5'-end which create a 30 restriction enzyme site, thus facilitating cloning. The sequence of ssc5h34A is

5' GAT CTC TAG AGA AAT CAA TAT GGT GAC AGA GTC A 3'.

The PCR reaction, which was initially described by Saiki et al. (1986), is carried out essentially as described in Lee et al. (1988), except that the template for the cDNA is RNA isolated from HCV infected chimpanzee liver, as described in Section IV.C.2., or from viral particles isolated from HCV infected chimpanzee serum, as described in Section IV.A.1. In addition, the annealing conditions are less stringent in the first round of amplification (0.6M NaCl, and 25°C), since the part of the primer which will anneal to the HCV sequence is only 9 nucleotides, and there could be mismatches. Moreover, if ssc5h34A is used, the additional sequences not derived from the HCV genome tend to destabilize the primer-template hybrid. After the first round of amplification, the annealing conditions can be more stringent (0.066M NaCl, and 32°C-37°C), since the amplified sequences now contain regions which are complementary to, or duplicates of the primers. In addition, the first 10 cycles of amplification are run with Klenow enzyme I, under appropriate PCR conditions for that enzyme. After the completion of these cycles, the samples are extracted, and run with Taq polymerase, according to kit directions, as furnished by Cetus/Perkin-Elmer.

After the amplification, the amplified HCV cDNA sequences are detected by hybridization using a probe derived from clone 5h. This probe is derived from sequences upstream of those used to derive the primer, and does not overlap the sequences of the clone 5h derived primers. The sequence of the probe is

5' CCC AGC GGC GTA CGC GCT GGA CAC GGA GGT GGC CGC GTC
GTG TGG CGG TGT TGT TCT CGT CGG GTT GAT GGC GC 3'.

IV.N.1. Creation of HCV cDNA Library from liver of a Chimpanzee with infectious NANBH

An HCV cDNA library was created from liver from the chimpanzee from which the HCV cDNA library in Section IV.A.1. was created. The technique for creating the library was similar to that in Section IV.A.24, except for this different source of the RNA, and that a primer based on the sequence of HCV cDNA in clone 11b was used. The sequence of the primer was

5' CTG GCT TGA AGA ATC 3'.

IV.N.2. Isolation and nucleotide sequence of overlapping HCV cDNA in clone k9-1 to cDNA in clone 11b

Clone k9-1 was isolated from the HCV cDNA library created from the liver of an NANBH infected chimpanzee, as described in Section IV.A.25. The library was screened for clones which overlap the sequence in clone 11b, by using a clone which overlaps clone 11b at the 5'-terminus, clone 11e. The sequence of clone 11b is shown in Fig. 23. Positive clones were isolated with a frequency of 1 in 500,000. One isolated clone, k9-1, was subjected to further study. The overlapping nature of the HCV cDNA in clone k9-1, to the 5'-end of the HCV-cDNA sequence in Fig. 26 was confirmed by probing the clone with clone Alex 46; this latter clone contains an HCV cDNA sequence of 30 base pairs which corresponds to those base pairs at the 5'terminus of the HCV cDNA in clone 14i, described supra..

The nucleotide sequence of the HCV cDNA isolated from clone k9-1 was determined using the techniques described supra. The sequence of the HCV cDNA in clone k9-1, the overlap with the HCV cDNA in Fig. 26, and the

amino acids encoded therein are shown in Fig. 46.

The HCV cDNA sequence in clone k9-1 has been aligned with those of the clones described in Section IV.A.19. to create a composite HCV cDNA sequence, with the k9-1 sequence being placed upstream of the sequence shown in Fig. 32. The composite HCV cDNA which includes the k9-1 sequence and the amino acids encoded therein is shown in Fig. 47.

The sequence of the amino acids encoded in the 5'-region of HCV cDNA shown in Fig. 47 has been compared with the corresponding region of one of the strains of Dengue virus, described supra., with respect to the profile of regions of hydrophobicity and hydrophilicity. This comparison showed that the polypeptides from HCV and Dengue encoded in this region, which corresponds to the region encoding NS1 (or a portion thereof), have a similar hydrophobic/hydrophilic profile.

The information provided infra. allows the identification of HCV strains. The isolation and characterization of other HCV strains may be accomplished by isolating the nucleic acids from body components which contain viral particles, creating cDNA libraries using polynucleotide probes based on the HCV cDNA probes described infra., screening the libraries for clones containing HCV cDNA sequences described infra., and comparing the HCV cDNAs from the new isolates with the cDNAs described infra. The polypeptides encoded therein, or in the viral genome, may be monitored for immunological cross-reactivity utilising the polypeptides and antibodies described supra. Strains which fit within the parameters of HCV, as described in the Definitions section, supra., are readily identifiable. Other methods for identifying HCV strains will be obvious to those of skill in the art, based upon the information provided herein.

Industrial Applicability

The invention, in the various manifestations disclosed herein, has many industrial uses, some of which are the following. The HCV cDNAs may be used for the

5 design of probes for the detection of HCV nucleic acids in samples. The probes derived from the cDNAs may be used to detect HCV nucleic acids in, for example, chemical synthetic reactions. They may also be used in screening programs for anti-viral agents, to determine the effect of

10 10 the agents in inhibiting viral replication in cell culture systems, and animal model systems. The HCV polynucleotide probes are also useful in detecting viral nucleic acids in humans, and thus, may serve as a basis for diagnosis of HCV infections in humans.

15 In addition to the above, the cDNAs provided herein provide information and a means for synthesizing polypeptides containing epitopes of HCV. These polypeptides are useful in detecting antibodies to HCV antigens. A series of immunoassays for HCV infection,

20 20 based on recombinant polypeptides containing HCV epitopes are described herein, and will find commercial use in diagnosing HCV induced NANBH, in screening blood bank donors for HCV-caused infectious hepatitis, and also for detecting contaminated blood from infectious blood donors.

25 The viral antigens will also have utility in monitoring the efficacy of anti-viral agents in animal model systems. In addition, the polypeptides derived from the HCV cDNAs disclosed herein will have utility as vaccines for treatment of HCV infections.

30 The polypeptides derived from the HCV cDNAs, besides the above stated uses, are also useful for raising anti-HCV antibodies. Thus, they may be used in anti-HCV vaccines. However, the antibodies produced as a result of immunization with the HCV polypeptides are also useful in

35 35 detecting the presence of viral antigens in samples. Thus,

- they may be used to assay the production of HCV polypeptides in chemical systems. The anti-HCV antibodies may also be used to monitor the efficacy of anti-viral agents in screening programs where these agents are tested
- 5 in tissue culture systems. They may also be used for passive immunotherapy, and to diagnose HCV caused NANBH by allowing the detection of viral antigen(s) in both blood donors and recipients. Another important use for anti-HCV antibodies is in affinity chromatography for the
- 10 purification of virus and viral polypeptides. The purified virus and viral polypeptide preparations may be used in vaccines. However, the purified virus may also be useful for the development of cell culture systems in which HCV replicates.
- 15 Cell culture systems containing HCV infected cells will have many uses. They can be used for the relatively large scale production of HCV, which is normally a low titer virus. These systems will also be useful for an elucidation of the molecular biology of the
- 20 virus, and lead to the development of anti-viral agents. The cell culture systems will also be useful in screening for the efficacy of antiviral agents. In addition, HCV permissive cell culture systems are useful for the production of attenuated strains of HCV.
- 25 For convenience, the anti-HCV antibodies and HCV polypeptides, whether natural or recombinant, may be packaged into kits.
- The method used for isolating HCV cDNA, which is comprised of preparing a cDNA library derived from
- 30 infected tissue of an individual, in an expression vector, and selecting clones which produce the expression products which react immunologically with antibodies in antibody-containing body components from other infected individuals and not from non-infected individuals, may also be
- 35 applicable to the isolation of cDNAs derived from other

heretofore uncharacterized disease-associated agents which
are comprised of a genomic component. This, in turn,
could lead to isolation and characterization of these
agents, and to diagnostic reagents and vaccines for these
5 other disease-associated agents.

10

15

20

25

30

35

CLAIMS

1. A purified HCV polynucleotide.
- 5 2. A recombinant HCV polynucleotide.
3. A recombinant polynucleotide comprising a sequence derived from an HCV genome or from HCV cDNA.
- 10 4. A recombinant polynucleotide encoding an epitope of HCV.
5. A recombinant vector containing the polynucleotide of claim 2, or claim 3, or claim 4.
- 15 6. A host cell transformed with the vector of claim 5.
7. A recombinant expression system comprising an open reading frame (ORF) of DNA derived from an HCV genome or from HCV cDNA, wherein the ORF is operably linked to a control sequence compatible with a desired host.
- 25 8. A cell transformed with the recombinant expression system of claim 7.
9. A polypeptide produced by the cell of claim 8.
- 30 10. Purified HCV.
11. A preparation of polypeptides from the HCV of claim 10.

35

12. A purified HCV polypeptide.

13. A purified polypeptide comprising an epitope which is immunologically identifiable with an 5 epitope contained in HCV.

14. A recombinant HCV polypeptide.

15. A recombinant polypeptide comprised of a 10 sequence derived from an HCV genome or from HCV cDNA.

16. A recombinant polypeptide comprised of an HCV epitope.

15 17. A fusion polypeptide comprised of an HCV polypeptide.

18. A monoclonal antibody directed against an HCV epitope.

20

19. A purified preparation of polyclonal antibodies directed against HCV.

25 20. A particle which is immunogenic against HCV infection comprising a non-HCV polypeptide having an amino acid sequence capable of forming a particle when said sequence is produced in a eukaryotic host, and an HCV epitope.

30

21. A polynucleotide probe for HCV.

35

22. A kit for analyzing samples for the presence of polynucleotides derived from HCV comprising a polynucleotide probe containing a nucleotide sequence from

HCV of about 8 or more nucleotides, in a suitable container.

23. A kit for analyzing samples for the presence of an HCV antigen comprising an antibody directed against the HCV antigen to be detected, in a suitable container.

24. A kit for analyzing samples for the presence of an antibodies directed against an HCV antigen comprising a polypeptide containing an HCV epitope present in the HCV antigen, in a suitable container.

25. A polypeptide comprised of an HCV epitope, attached to a solid substrate.

26. An antibody to an HCV epitope, attached to a solid substrate.

27. A method for producing a polypeptide containing an HCV epitope comprising incubating host cells transformed with an expression vector containing a sequence encoding a polypeptide containing an HCV epitope under conditions which allow expression of said polypeptide.

28. A polypeptide containing an HCV epitope produced by the method of claim 27.

29. A method for detecting HCV nucleic acids in a sample comprising:

(a) reacting nucleic acids of the sample with a probe for an HCV polynucleotide under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample; and

(b) detecting a polynucleotide duplex which contains the probe.

30. An immunoassay for detecting an HCV antigen
5 comprising:

(a) incubating a sample suspected of containing an HCV antigen with a probe antibody directed against the HCV antigen to be detected under conditions which allow the formation of an antigen-antibody complex; and

10 (b) detecting an antigen-antibody complex containing the probe antibody.

31. An immunoassay for detecting antibodies directed against an HCV antigen comprising:

15 (a) incubating a sample suspected of containing anti-HCV antibodies with a probe polypeptide which contains an epitope of the HCV, under conditions which allow the formation of an antibody-antigen complex; and

20 (b) detecting the antibody-antigen complex containing the probe antigen.

32. A vaccine for treatment of HCV infection comprising an immunogenic polypeptide containing an HCV epitope wherein the immunogenic polypeptide is present in
25 a pharmacologically effective dose in a pharmaceutically acceptable excipient.

33. A vaccine for treatment of HCV infection comprising inactivated HCV in a pharmacologically
30 effective dose in a pharmaceutically acceptable excipient.

34. A vaccine for treatment of HCV infection comprising attenuated HCV in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

35. A tissue culture grown cell infected with HCV.

36. The HCV infected cell of claim 35, wherein
5 the cell is of a human macrophage cell line, or is of a hepatocyte cell line, or is of a mosquito cell line, or is of a tick cell line, or is of a mouse macrophage cell line, or is an embryonic cell.

10 37. The HCV infected cell of claim 35, wherein
the cell is of a cell line derived from liver of an HCV
infected individual.

38. A method for producing antibodies to HCV
15 comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope in an amount sufficient to produce an immune response.

39. A method for producing antibodies to HCV
20 comprising administering to an individual the polypeptide preparation of claim 11, wherein the preparation contains at least 1 immunogenic polypeptide, and the administering is of an amount sufficient to produce an immune response.

25 40. A method for isolating cDNA derived from the genome of an unidentified infectious agent,
comprising:

(a) providing host cells transformed with expression vectors containing a cDNA library prepared from
30 nucleic acids isolated from tissue infected with the agent and growing said host cells under conditions which allow expression of polypeptide(s) encoded in the cDNA;

(b) interacting the expression products of the cDNA with an antibody containing body component of an
35 individual infected with said infectious agent under

conditions which allow an immunoreaction, and detecting antibody-antigen complexes formed as a result of the interacting;

- (c) growing host cells which express polypeptides that form antibody-antigen complexes in step (b) under conditions which allow their growth as individual clones and isolating said clones;
- (d) growing cells from the clones of (c) under conditions which allow expression of polypeptide(s) encoded within the cDNA, and interacting the expression products with antibody containing body components of individuals other than the individual in step (a) who are infected with the infectious agent and with control individuals uninfected with the agent, and detecting antibody-antigen complexes formed as a result of the interacting;
- (e) growing host cells which express polypeptides that form antibody-antigen complexes with antibody containing body components of infected individuals and individuals suspected of being infected, and not with said components of control individuals, under conditions which allow their growth as individual clones and isolating said clones; and
- (f) isolating the cDNA from the host cell clones of (e).

30

35